SEP 21 1922
UNIV. OF MICH.

The Journal of Educational Psychology

Devoted Primarily to the Scientific Study of Problems of Learning and Teaching.

BOARD OF EDITORS:

HAROLD ORDWAY RUGG, *Chairman.*
Lincoln School of Teachers College.
Teachers College, Columbia University.

JAMES CARLETON BELL,
Brooklyn Training School for Teachers.

FRANK NUGENT FREEMAN,
University of Chicago.

ARTHUR IRVING GATES,
Teachers College, Columbia University.

VIVIAN ALLEN CHARLES HENMON,
University of Wisconsin.

RUDOLF PINTNER,
Teachers College, Columbia University.

BEARDSLEY RUML,
Carnegie Corporation, New York City.

LEWIS MADISON TERMAN,
Leland Stanford University.

EDWARD LEE THORNDIKE,
Teachers College, Columbia University.

LAURA ZIRBES, *Assistant Editor.*
Lincoln School of Teachers College.

VOL. XIII.
No. 6.

SEPTEMBER, 1922

$4.00 a year.
60c. per copy.

CONTENTS

AN EXPERIMENT IN LEARNING AN ABSTRACT SUBJECT. *Edward L. Thorndike and C. B. Upton* 321

A CLASS EXPERIMENT IN LEARNING. *Walter F. Dearborn and Edward A. Lincoln* 330

LANGUAGE ERROR TESTS. *G. M. Wilson* 341

SOME EVIDENCE OF AN ADOLESCENT INCREASE IN THE RATE OF MENTAL GROWTH. *Katherine Murdock and Louis R. Sullivan* 350

TENTATIVE ORDER OF DIFFICULTY OF THE TERMAN VOCABULARY WITH VERY YOUNG CHILDREN. *Margaret V. Cobb* 357

SOME RETESTS WITH THE STANFORD-BINET SCALE. *Kate Gordon* 363

"THE CONSTANCY OF THE IQ" AGAIN. *Florence M. Teagarten* . . 366

NOTES ON ARTICLES IN EDUCATIONAL PSYCHOLOGY IN CURRENT ISSUES OF OTHER MAGAZINES 373

NEW PUBLICATIONS IN EDUCATIONAL PSYCHOLOGY AND RELATED FIELDS OF EDUCATION 378

Published Monthly Except June to August by
WARWICK and YORK, Inc.,
York, Pa. *Baltimore, Md.*

Entered as Second Class matter November 15, 1921, at the Post Office at York, Pennsylvania, under the Act of March 3, 1879.

The Journal of Educational Psychology

is published monthly, by Warwick & York, Inc., Baltimore, Md., and York, Pa. The issues for 1922 make Volume XIII. Title page and index are bound in December number.

Manuscripts for publication, books or other materials for review, and news items should be addressed to Harold Ordway Rugg, Chairman of the Editorial Board, 59 Edgecliffe Terrace, Yonkers, New York.

The price of the Journal is $4.00 a year, payable in advance. Foreign postage is 30 cents extra. Single issues are 60 cents each, and less than a full year's subscription is charged at single issue rate for every month ordered.

Subscribers should notify the publishers of change in address at least four weeks before publication of the issue with which change is to take effect. No claim for non-receipt of an issue will be entertained unless made within two weeks after receipt of the next succeeding issue.

Warwick & York, Inc.,
Baltimore, Md.

TEACHERS' HEADQUARTERS

European
Plan

Entirely
Fire-Proof

Centrally
Located

EDWARD DAVIS
Manager

HOTEL RENNERT
BALTIMORE, MARYLAND

RECENT PUBLICATIONS

Brief Introductory Psychology for Teachers.

By Edward K. Strong, Jr.

xii, 241 *pages.* $1.60. *Postage* 10¢.

The Output of Professional Schools for Teachers.

By Charles E. Benson.

x, 88 *pages.* $1.50. *Postage* 10¢.

Textbook Selection.

By R. H. Frazen and F. B. Knight.
Introduction by Ernest Horn.

94 *pages.* $1.20. *Postage* 6¢.

Warwick & York, Inc., Publishers, Baltimore.

THE JOURNAL OF
EDUCATIONAL PSYCHOLOGY

Vol. XIII September, 1922 No. 6

AN EXPERIMENT IN LEARNING AN ABSTRACT SUBJECT[1]

EDWARD L. THORNDIKE AND C. B. UPTON

Institute of Educational Research, Teachers College, Columbia University

The purpose of the experiment was to observe the process of learning an abstract subject by adult learners in a case where the task is for them of somewhat the same novelty and difficulty as the learning of algebra is for the first-year high school pupil or as the learning of physics is for the third year pupil. We were also concerned with finding out how profitable it seemed to be for teachers to study themselves as learners in the case of such abstract material.

The subjects were a score of college graduates—teachers of mathematics. The task was to acquire understanding of certain elementary laws of electricity and magnetism. It was defined by the following instructions:

EXPERIMENT, APRIL, 1922

LEARNING AN ORGANIZED ABSTRACT SUBJECT

Study pages 1 to 22 (up to Section 15) of FRANKLIN and ESTY, "Elements of Electrical Engineering; Direct Currents" for 6 hours. Keep notes of difficulties and of how you met them.

There are six or seven copies on reference in the library.

[1] The investigation reported in this article was aided by a grant from the Commonwealth Fund.

Begin on page 1; refer back to pages XI, XII and XIII as is needed. Refer to attached notes on physics as is needed.[1] Refer to physics text-book if you find the attached notes insufficient. There are several copies on reference in the library.

Then work problems 4 to 18 inclusive on pages 448 to 453, so far as you have time in 4 hours more. Refer back to pages 1 to 22 as you need to. Keep notes as before.

The answers which the text-book provides were erased, since it seemed more instructive to make the first experiment without aid

---

[1] The attached notes were as follows:

C.G.S. Units, Which May Be Needed, in Solving the Examples Assigned

Dyne	= force which acting on 1 gr. for 1 second increases its velocity 1 cm.
The number of dynes	= weight in grams times acceleration of gravity in centimeters per second per second.
Erg	= unit of work or energy = 1 dyne acting through a distance of 1 cm.
Joule	= 10^7 ergs.

English Units

Foot-pound	= force required to move 1 lb. a distance of 1 ft.
Pound-inch	= force required to move 1 lb. 1 in.

Definitions

Torque equals the force times the distance through which the power works.

Center of gravity is the point on which any object will balance.

Armature, a core of metal surrounded by a coil of wire, rotating near the poles of magnets in a dynamo.

Dynamo, a machine to convert mechanical energy expended upon it into electrical energy.

Equivalent units of measure that may be needed in solving the examples assigned:

1 horse power	=	33,000 ft.-lbs. per minute
1 horse power	=	746 watts
1 watt	=	0.00134 horse power
1 joule	=	0.74 ft.-lb.
1 watt	=	44.3 ft.-lb. per minute
1 cm.	=	0.3937 in.
1 ft.	=	30.48 cm.
1 lb.	=	453.6 gr.
1 gram	=	0.0353 oz. avoirdupois
1 kilogram	=	2.2046 lb. avoirdupois

from them. We hope to repeat the experiment with a similar group using the answers.

It was explained to most of the subjects that the 6 hours of work on the text, the 4 hours of work on the text, and the problems could be divided up and alternated in any way that the subject might choose. All work was signed with a number or pseudonym. The subjects were asked to hand in all their work and to put down in writing any matters concerning the learning process, such as difficulties and the means taken to overcome them which seemed instructive, taking out time therefor.

This last injunction resulted in only rather meager notes, probably for two reasons. First, the teacher does not, unless he has an exceptional interest in the psychology of learning, watch himself learn when he learns. On the contrary he becomes absorbed in the content which he is trying to master. In the second place, there is a natural tendency to avoid the labor of making such notes.

The 10 hours were spent at times of the subject's choice during a week. At the close of the week certain questions about their experiences were asked of all the subjects and a 12-minute test was given. What follows in this article is the writer's opinion based on the answers to these questions, the notes on learning, the written work, and the test.

The experiment seems a useful one as a means of increasing the subject's appreciation of pupils' difficulties and sympathy with their efforts, and as a means of increasing the experimenter's knowledge of learning. The task assigned is somewhere nearly as hard for college graduates who know physics only from an elementary course, as beginning algebra and physics and economics are for their students. This is evidenced by the records with the problems. The highest score was 12½ correct out of 15; the lowest was 0; the median was 5; half were from 3 to 7 inclusive. The errors gave by their nature, and especially their variety, much the same impression of inadequacy, confusion and carelessness that one gets from the results of a hard assignment to a high school class in an abstract subject.

It is not grossly inaccurate to say to a group of teachers who have done this experiment: "This is to you approximately what a hard series of lessons is to a high school class. The difficulties and discouragements and confusion which you have felt are what they feel. Although you did your best, your work looks stupid and careless, much as theirs does. It looks in places as if you 'did not try' or

'would not make yourself think,' or 'did not keep your attention on what you were doing.'" The experiment should be a warning and protection against underestimating pupils' difficulties and imputing to perverseness or lack of effort results which are really due to the general laws of mental action. It should enforce the general lesson of psychology that commands and exhortations and rebukes are only a very small part of teaching, the great part being to discern the forces of the pupils' own minds and hearts and manœuver them to desirable ends.

So much for the benefits to be derived from the experiment in making teachers more humane and reasonable. With respect to our knowledge of the learning process itself, its chief result is its evidence that gifted, highly trained adults are much more like ordinary untrained children in methods and procedure in learning than it has been the fashion to suppose. It has been customary to contrast the inferior young, untrained child sharply with the superior, trained adult in that the former is impulsive, leaping before he looks; is uncritical, accepting any idea that happens to strike him regardless of its appropriateness; and is confused and dreamy and irrelevant, lacking clear-cut ideas of what he is to think and why he is to think it. This contrast is treated as general, pervading all the thinking of the two types of mind, and constitutional, depending on their distinctive natures, not on the tasks in which they engage. The statement that superior trained adults seem more rational, critical and clear-headed in large measure because they do only what is easy for them, would be regarded by all educational psychologists as a paradox and by most of them as a very silly one. Yet our inspection of the work of this score of superior trained adult thinkers strongly suggests that the paradox has a large element of truth, that the difference in thinking is not all-pervasive and constitutional, but is, in part, a consequence of specialized habits.

Consider, for example, the answers given by these gifted trained adults to the question "Why are magnets which are used to pick up pieces of iron bent into U form?" which was asked at the end of the 10 hours of study of the elementary principles of magnetism. Many of them show an impulsive, uncritical, and confused thinking to a notable degree—to the person who is really master of the topic.

> **(*A*) The intensity of magnetic field is increased by bringing poles near each other. Attraction between two poles is product of their intensities divided by distance squared, and as distance is decreased the intensity increases inversely as its square.**

(*B*) Magnetic force tends to concentrate at one spot or another. Therefore, if the iron is bent horseshoe fashion and magnetized, there is more strength.

(*C*) To get both poles on a level to attract either or both poles of the article you want to pick up.

(*D*) In order that the object may lie along the lines of intensest force, thus the strongest pull being conducted through the material picked up.

(*E*) Magnets are bent into U form in order to create a magnetic field of .....

(*F*) Magnets are bent into U form to increase intensity of field. The two poles being .....

(*G*) So that the poles are more convenient for use. More handy for use than a long magnet.

(*H*) The magnet is in U form to bring poles nearer together.

(*I*) To strengthen field; $F = \frac{m'm''}{r^2}$ as distance r decreases, F increases.

(*J*) Pull is both ways and stronger north and south poles give equal pull.

(*K*) Bringing both poles together condenses the external magnetic field, thus making it stronger. Both poles are brought to bear on an object, thus .....

(*L*) To attract both north and south pole. The flux is from north to south pole; by having U-shaped magnet the two poles are near enough to make the circuit.

(*M*) So that both poles may operate at once; putting the poles together concentrates the field, *i.e.*, shortens and pushes closer the lines of force.

(*N*) Because the closer the poles are together the more intense is the magnetic flux.

(*O*) Magnets are bent into U form to bring the poles nearer together.

(*P*) Bent to give two poles so that the current is attracted to the second pole, otherwise there would be no flow of current.

(*Q*) U form is due to the fact that like poles repel and unlike poles attract. If the poles are close together the force of attraction is greater and exerts greater force on external objects.

(*R*) The poles are opposite in attraction and have power that is concentrated.

(*S*) So that the entire force of the magnet is at one point and so that both directions of force may be exerted at once.

(*T*) Double the strength, or attraction on the iron.

Impulsiveness and uncriticalness are witnessed further by the large number of wrong answers in comparison with the number of omissions or confessions of inability. In the book problems there were (counting each task requiring a separate answer as one problem) 341 answers given of which only 53 per cent were right. In the case

of few of these was there any apology or other expression of doubt about the answer. In the test problems a similar state of affairs obtains. The extreme of uncriticalness appears where the thinker merely "fishes about" for any promising formula and puts the number of the problem into that formula, almost haphazard. Consider these answers all given to the same problem and all wrong.

15,503 lines
15.5035
93,020 lines
348,000 cos 63
174,000 cos 63
4.739 gauss
78,994 dynes
7.8996

A general mental confusion is evidenced by much of the work, especially perhaps by the very frequent failure to define the answer numbers as dynes, gausses, maxwells, lines, centimeters, or whatever they should be, and by the frequent attachment of erroneous names to the answer numbers.

Thus in the test the question: "What is the intensity of the magnetic field due to a unit pole at a distance of .001 cm.?" received a correct number answer six times but the name was correct only once.

The "childishness" of these adults appears also in their eagerness to learn from some authority whether the answers which they obtain are right, rather than trust to their own proofs and checks.

We are not, of course, saying that the work of first-year high school pupils *at this task* would not be easily distinguished from these adults' work at it. Nor do we say that these adults are as childish at it as first-year pupils are at their tasks in mathematics, grammar, or science. What is claimed here is not that the difference between the thinking of superior trained adults and that of children is zero, but that it is less than the orthodox educational psychology and child psychology of the last two decades has taught. In proportion as a task for thought is novel and hard, the superior trained adult tends to jump at conclusions instead of planfully mastering each necessary step, to let results stand without surety that they are true and useful, and to become confused, making computations and statements without any clear realization of what he is doing and why he does it.[1]

[1] I may be permitted to note that I observed these tendencies in my own work at this task.

On the Psychology of Teaching in General

The experiment provides evidence corroborating certain pedagogical doctrines which are already accepted or on the road to acceptance. There was general agreement among the subjects that each item of fact or principle should be applied as soon as learned, that miscellaneous problems requiring the selection of appropriate principles should be given later, and that many short problems with a minimum of computation and interpretation of complex situations should be given first. There was general agreement that, even with the very scanty problem list of the book used, the problems explained the text as truly as the text explained the problems. There was substantial agreement that the learner should be enabled to ascertain whether his work was correct very soon after he finished any part of it and very often along its course. Sample problems requiring the use of formulas should be solved in the text as illustrations of the formulas.

The irritation caused by features of the task that were, or at least seemed, irrelevant, such as laborious computations, a problem about the torque on a pulley, and the recalculation of an answer in another system of units, was notable. Irritation at being thwarted by a difficulty that was, or at least seemed, not provided for by the text was also great. The text left, or seemed to leave, the meaning of gauss and the relation of a gauss to the force exerted by a unit pole at a distance of 1 cm. to be inferred. We estimate that if it had stated it clearly and emphatically there would have been a saving per person of 30 minutes time, much irritation and at least ten per cent in errors.[1]

Finally the facts of the experiment suggested to the writer as highly probable an important amendment to the common view of the difference between the "easy" subjects like English, History, French, or book-keeping and the "hard" subjects like mathematics, Latin, or physics. The common view is that the latter are hard because they are more abstract, more organized, more rigorous and more precise. They require, to a greater degree, analysis and reasoning in place of memory, systematized related knowledge in place of an accumulation of details, an absolute adherence to certain rules of the game rather than a general following of their spirit, and 100 per cent precision in the operation of the mental bonds instead of a certain free play for

[1] Neither this paragraph, nor anything else in this report, is intended or should be interpreted as an adverse criticism of the text-book pages used. Our experiment put them to a use very different from that for which they were designed.

errors of moderate amount. This is all true; but it may also be true that part of the greater difficulty is due to the sheer unfamiliarity of the date, the greater amount to be learned per unit of time, the more subtle symbolism, and the larger number of missing links—facts that have to be read between the lines, connections that have to be worked out by the pupil by inference. If high school pupils studied the history of philosophical systems or of credit facilities, history would be made harder by the unfamiliar data. If, when they had studied a certain phenomena, say the causes of revolt in Greece and Rome, they were expected to apply that knowledge to all cases of revolt anywhere, distinguishing when it was and when it was not applicable, history would be harder because learning the causes of revolt in Greece and Rome would then be learning a great deal more than a chronicle plus certain suggestions. If, in the text-book on history, e stood for envy, r for race antagonism, p for George Washington, the area of a square for the number of people concerned, l_1, l_2, l_3 etc., for certain typical cases previously taught, and the like, history would be harder. If any facts that a competent thinker could infer were omitted from statement in history (as if we read "Charles" instead of "King Charles," "Gregory" instead of "Pope Gregory." "He infringed upon the six most important civil rights of the population," instead of a long paragraph stating in detail what he did, history would be harder.

In the experiment all these four factors appeared: (1) Strength of pole, magnetic field, intensity of magnetic field, and magnetic flux are hard partly because they are strange to us. (2) Formulas like $F = mH$ or $\Phi = sH$ look like small bits of learning, but each is really, if understood, a large body of fact and a still larger possibility of handling other facts. Thirteen such formulas in 6 hours means a very large amount of learning per hour. (3) The symbolism which will in the end economize thought is in the beginning a burden. (4) Consider the number of inferences that a student must make to obtain a true and adequate learning of this single half page.

> "*Magnetic Flux.*—Consider a plane surface, s square centimeters in area, stretched across and at right angles to a uniform magnetic field of intensity H. The product sH is called the *magnetic flux* across the surface. That is:
>
> $$\phi = sH \qquad (4a)$$
>
> in which Φ is the magnetic flux across a plane surface of area s square centimeters at right angles to a uniform magnetic

field of intensity H. When the plane surface is not at right angles to the uniform magnetic field then:

$$\phi = sH \cos \theta \qquad (4b)$$

iu which θ is the angle between H and the normal to the surface. When the field is not uniform or when the surface is curved then:

$$\Delta\phi = H \cos \theta . \Delta s \qquad (4c)$$

in which $\Delta\phi$ is the flux across an element of the surface of which the area is Δs, H is the intensity of the field at the element of surface, and θ is the angle between H and the normal to the element of surface. In this case the total flux across a finite surface is found by integrating equation (4*c*) over the finite surface."

So, we repeat, the experiment suggests that these other characteristics deserve consideration along with abstractness, organization, rigor and precision. Precision and rigor may be rated as indispensable features of secondary education. Abstractness, organization, and subtle symbolisms are very desirable features of it for those whose intellects can handle them. What I have called *reading between the lines*, supplying necessary interpretations and connections, is desirable in moderation. The mere difference in amount of learning whereby an assignment in English or history seems more than it is and an assignment in mathematics or physics is much more than it seems, is not of any intrinsic educational value. The difference in strangeness is perhaps of negative value educationally, it being unwise to try to understand, analyze, compare and reason about facts until we have a certain familiarity with them as facts. In so far as Latin is hard because of the sheer strangeness of datives and ablatives, its hardness is chiefly an unfortunate interference to be overcome. In so far as algebra is hard because the use of letters to represent numbers is strange, the lesson to its teachers is to remove that much of its hardness as quickly as they can by suitable experiences with formulas and their evaluation.

A CLASS EXPERIMENT IN LEARNING

WALTER F. DEARBORN AND EDWARD A. LINCOLN

Psycho-educational Clinic, Graduate School of Education, Harvard University

In presenting to classes in Educational Psychology the facts about learning, the effects of practice and transfer of training it seems highly desirable to adopt a method which will furnish to the students evidence of an objective nature. The limited time at the disposal of the lecturer in the ordinary introductory course in this field precludes the possibility of adapting the typical experiments in learning to fit the occasion. The studies of Bryan and Harter, Swift, Book, and others have dealt with the learning of telegraphy, typewriting, or some other subject matter which required lengthy practice extending sometimes over months. In a previous article by one of the present writers a preliminary statement of the problem was made and suggestions were offered for experiments which seemed to meet the requirements of the situation.[1] Several years later there was published a more elaborate scheme for carrying out a class experiment in learning.[2] This experiment did not prove entirely satisfactory, but it has since been modified, after trial in a number of classes, and it now seems to serve its purpose excellently. It is the object of this paper to report the methods and results of the experiment as given in its final form.

The class in Educational Psychology was divided arbitrarily according to the seating arrangements into two groups of approximately equal size. On the first day of the experiment all members of the class were given three tests, hereafter designated as the Test Series. The first of these is called the Digit Letter Substitution Test. Sheets of paper were distributed, each containing at the top circles in which were inserted the numbers from 1 to 26. Each circle also contained one of the letters of the alphabet, the distribution being a random one. Under the code were printed a number of sentences, and the task consisted in translating letters into digits for a period of 5 minutes. As shown in Fig. 1 the arrangement of the sheet permits of rapid scoring. The subjects were told to work for accuracy rather

---

[1] Dearborn, W. F.: Experiments in Learning. *Journal of Educational Psychology*, Vol. I, pp. 373–384.

[2] Dearborn, W. F., and Brewer, John M.: Methods and Results of a Class Experiment in Learning. *Journal of Educational Psychology*, Vol. IX, No. 2, Feb., 1918, pp. 63–82.

O 1	V 2	C 3	F 4	I 5	K 6	B 7	M 8
T 9	X 10	E 11	W 12	N 13	H 14	G 15	L 16	D 17
J 18	S 19	R 20	U 21	Z 22	P 23	Y 24	A 25	Q 26

in the wind. The boys also	5	13	9	14	11	12	5	13	17	9	14	11	7	1	24	25	16	19	1	
spin tops, and they play	19	23	5	13	9	1	23	19	25	13	17	9	14	11	24	23	16	25	24	
" seesaw," and jump the rope.	19	11	11	19 25	25	12	25	13	17	18	21	8	23	9	14	11	20	1	23	11
Boys in Korea are fond of	7	1	24	19	5	13	6	1	20	11	25	25	20	11	4	1	13	17	1	4
fishing. Nearly every																				
boy has a fishing rod and																				
goes fishing whenever																				
he can. Sometimes the																				

FIG. 1.—Sample of digit-letter substitution test.

than speed, and to correct mistakes, and no account was taken of the errors in making up the scores.[1]

The second test is designated as the Complex Dotting Test. For

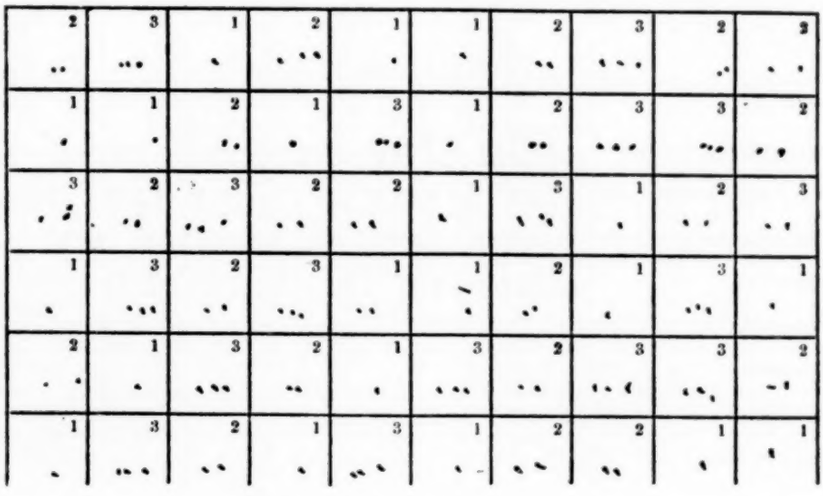

FIG. 2—Sample of dotting test.

[1] In one of the previous experiments the curve of errors was plotted, and it was shown that the errors might be disregarded.

this test sheets of paper are ruled into half-inch squares, each of which contains either the number 1, 2 or 3 (See Fig. 2). The numbers are distributed at random throughout the 150 squares on the sheet. The task is to put in each square as rapidly as possible the number of dots called for by the number in that square. One minute is allowed, and the score is the number of squares completed.

Test 3 is a modification of the code test of the Stanford-Binet Scale.[1] This code is reproduced in Fig. 3, and Fig. 4 shows part of a test sheet. Five minutes are allowed for this test, and the score is the number of letters translated. As in the other tests, no allowance is made for errors.

After these tests had been given the class was practiced for 10 minutes in the substitution of shorthand symbols for about a hundred of the most common words and word phrases in the English language. These symbols were printed on a key, a copy of which was given to each subject (Fig. 5). The material in which the substitutions were made was furnished by double spacing a book which was being printed at the University Press at the time and striking off sufficient copies on inexpensive paper. Figure 6 shows a section of this practice material in which the symbols have been substituted.

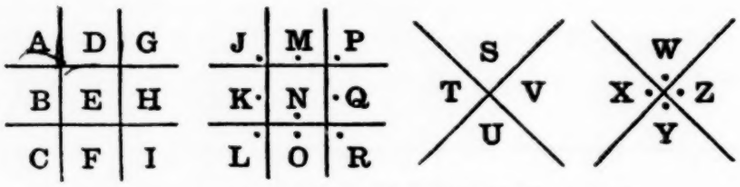

FIG. 3.—Code for code substitution test.

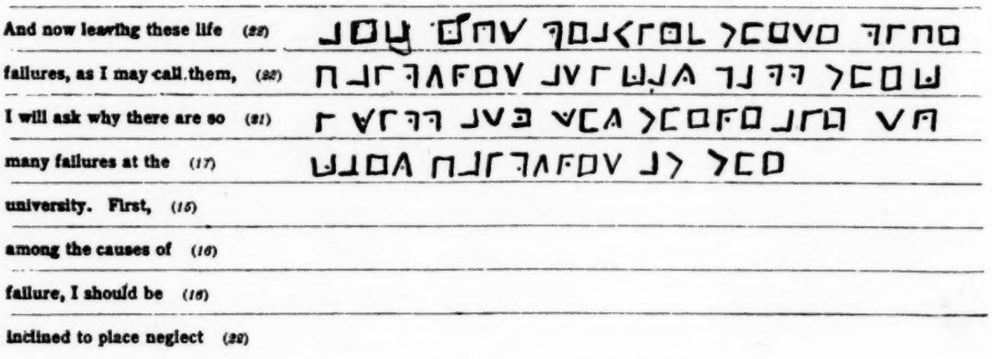
And now leaving these life (22)
failures, as I may call them, (22)
I will ask why there are so (21)
many failures at the (17)
university. First, (15)
among the causes of (16)
failure, I should be (15)
inclined to place neglect (22)

FIG. 4.—Sample of code substitution test.

[1] This code test seems to have first been described by Healy and Fernald in *Psychological Review Monographs*, Test No. XI, Vol. XIII, No. 2.

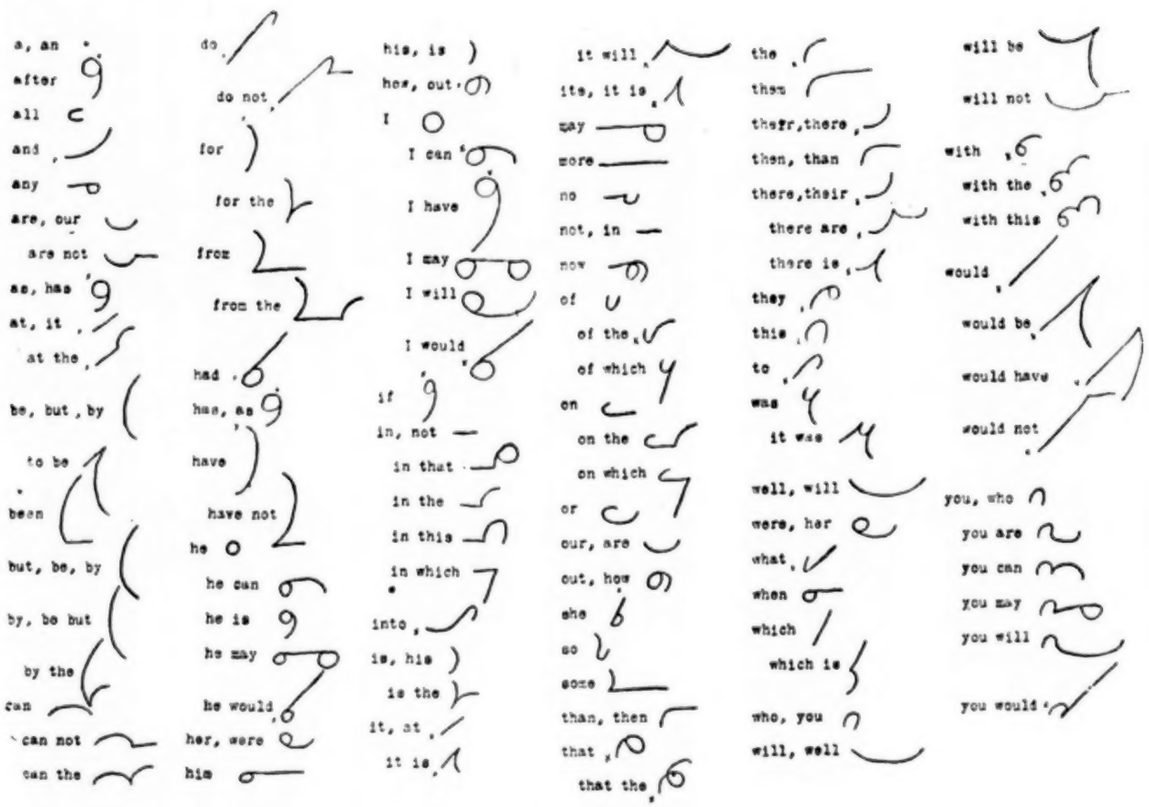

FIG. 5.—Shorthand key.

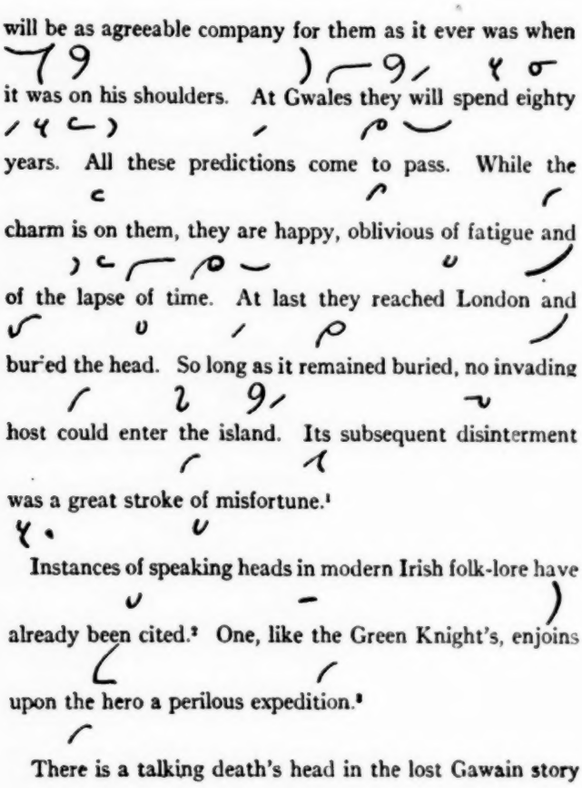
will be as agreeable company for them as it ever was when
it was on his shoulders. At Gwales they will spend eighty
years. All these predictions come to pass. While the
charm is on them, they are happy, oblivious of fatigue and
of the lapse of time. At last they reached London and
bur'ed the head. So long as it remained buried, no invading
host could enter the island. Its subsequent disinterment
was a great stroke of misfortune.[1]
Instances of speaking heads in modern Irish folk-lore have
already been cited.[2] One, like the Green Knight's, enjoins
upon the hero a perilous expedition.[3]
There is a talking death's head in the lost Gawain story

FIG. 6.—Sample of the practice material showing the shorthand substitutions.

During the next 15 days, exclusive of Sundays, this practice continued. One group of the class practiced one 20-minute period a day, and the other group practiced 20 minutes in two 10-minute periods, one in the morning and the other in the evening. Each man counted and recorded the number of substitutions made in each period, also noting the mental and physical conditions under which he did his work. At the end of the experiment each man tabulated his results and plotted his own curve of learning.

On the fifth and tenth days of the practice period check tests

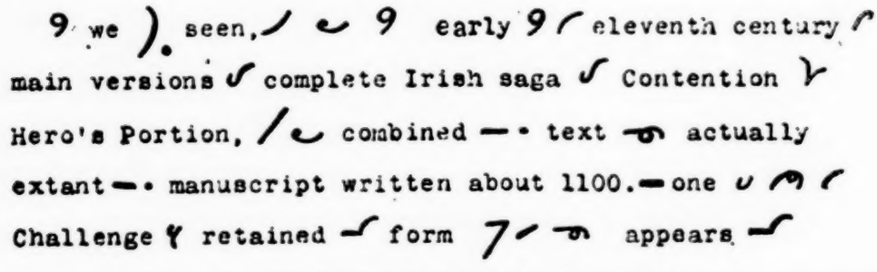
we seen, early eleventh century
main versions complete Irish saga Contention
Hero's Portion, combined text actually
extant manuscript written about 1100. one
Challenge retained form appears

FIG. 7.—Sample of check test material.

were given in the class. These tests reversed the operations of the practice. In a mimeographed passage from the same book which was used as practice material, shorthand symbols had been substituted in place of the words and phrases which appeared in the key, and under each symbol the proper word or phrase was written by the subject. Four minutes were allowed for each of these tests. In Fig. 7 a section of one of these tests is reproduced.

The final period of practice was held in the class room after the students had been working individually for 14 days.[1] At this time the test series was repeated.

As it was desired to show the ideal procedure for a transfer experiment the test series was given to a control group on the same days in which the practiced group took their first and last tests. This control group was composed of a class in Educational Psychology at Radcliffe College and a class in the History of Education at Harvard. Each class contained both graduate and undergraduate students. The following diagram, patterned after one in the previous article, illustrates the general plan of the experiment.

AN EXPERIMENT IN LEARNING

PRACTICED GROUP		
1st day	2nd, 3rd.....15th day	16th day
3 tests and first practice given	Practice series with shorthand symbols	Last practice given; 3 tests repeated.
UNPRACTICED OR CONTROL GROUP		
1st day	2nd, 3rd.....15th day	16th day
3 tests given	No intervening practice	3 tests repeated

After each student had made his own table of results and had plotted his curve of learning the records were collected and turned over to various groups in the class for the working out of certain problems which were made the subject of class reports. The findings presented in the more important of these reports are set forth in the following paragraphs.

Perhaps the greatest significance of the results of this experiment lies in the light they shed upon the problem of the transfer of training. The average scores of both practiced and control groups in the initial and final tests are set down in Table 1, together with the percentages of improvement shown in the test and practice series.

[1] The length of the practice may be changed to suit the conditions of the course in which the experiment is used.

TABLE I.—SCORES AND PERCENTAGES OF IMPROVEMENT OF PRACTICED AND CONTROL GROUPS

	Average initial score	Average final score	Per cent improvement
I. Practiced group:			
(*a*) Shorthand Substitution	72	467	548
(*b*) Digit letter Substitution	98	116	18
(*c*) Dotting	102	118	16
(*d*) Code Substitution	103	118	15
II. Control group:			
(*a*) Shorthand Substitution	No practice given.		
(*b*) Digit letter Substitution	93	103	11
(*c*) Dotting	102	113	11
(*d*) Code Substitution	92	116	26

There was plainly a very great amount of improvement in the ability to make the shorthand substitutions, as the class made an average gain of 548 per cent. The gains in the test series, however, were much smaller, namely, 15, 16, and 18 per cent. These test series gains become even smaller when the results of the control group have been taken into consideration. In the Digit Letter Substitution the mere repeating of the test in the control group shows an 11 per cent improvement. As the practiced group only improved 18 per cent in this test there is left a net improvement of 18 minus 11 or 7 per cent which may be attributed to transfer. In the same way we find a 5 per cent improvement which may be attributed to transfer in the Complex Dotting test, and in the Code Substitution there is no transfer or facilitation at all, but rather interference, as the control group improves 26 per cent while 15 per cent is all that is found in the practiced group.

In comparison with the improvement of 548 per cent which took place as a result of the practice it seems fair to conclude that the transfer was insignificant. We must, however, enter into another phase of the problem. Before we can decide on the significance of the above percentages we must know something about the zero points and the physiological limits in the various tests. Fortunately, the Digit Letter and Code Substitution Tests had each been used for practice material in previous experiments. In 12 days of practice

a group of students improved 463 per cent in the Digit Letter Test, and another class improved 537 per cent in Code Substitution.[1] Thus it seems that the possibilities of improvement in these tests were considerable. The fact that so little of this possible improvement appears in one test and that there is not an improvement but a loss in the other leads us to the conclusion that the skill acquired in one field was not transferable to another field, even though the second was very similar to the first. The possibility of one sort of learning interfering with another is also neatly illustrated.

The experiment also furnishes excellent material for the study of the effects of practice on the variability of a group. To get data on this point the average deviation of each group was calculated every third day throughout the experiment. These average deviations, together with the averages and coefficients of variability on the same days are tabulated in Table 2. On Plates I and II the deviations are shown by vertical lines.

TABLE II.—AVERAGES AND AVERAGE DEVIATIONS IN SHORTHAND SUBSTITUTION PRACTICE

Day	Group I		Coefficient of variability	Group II		Coefficient of variability
1	Average	150	0.08	Average	168	0.12
	Average deviation	12		Average deviation	20	
4	Average	344	0.13	Average	360	0.09
	Average deviation	46		Average deviation	34	
7	Average	535	0.12	Average	507	0.11
	Average deviation	62		Average deviation	57	
10	Average	665	0.13	Average	636	0.15
	Average deviation	88		Average deviation	84	
13	Average	844	0.11	Average	814	0.11
	Average deviation	90		Average deviation	90	
16	Average	960	0.11	Average	1001	0.11
	Average deviation	89		Average deviation	106	

The figures indicate that the heterogeneity of both groups was much increased by the practice. In Group 1 the average deviation on the first day was only 12, while on the last day it was 89. In Group II there was a rise from 20 to 106 in the average deviation.

[1] See above reference.

Manifestly the ability was more widely distributed at the end of the practice than it was at the beginning. This is shown in a rather striking way on the plates by the lines showing the daily individual performances of the slowest and most rapid learners in each group. When this change is expressed in terms of the coefficient of variability, it is seen that the increase in variability bears a fairly constant relation to the increase in the average skill of the groups.

The division of the class into two groups was for the purpose of investigating the effects of different distributions of practice time. The results of this study are best shown in Plate III where the daily averages of the two groups are plotted. These averages are tabulated in Table III. It will be seen that there are not large differences in the two curves, but the weight of evidence seems slightly in favor of the two short periods a day rather than one long one. The Group II curve is very smooth, while the daily performances of Group I

TABLE III.—DAILY AVERAGE SCORES OF TWO PRACTICED GROUPS

Day	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
Group I	150	223	254	344	390	462	535	560	602	665	793	786	844	829	936	960
Group II	168	247	304	360	414	446	507	545	588	636	686	740	814	859	908	1001

show more fluctuation, especially from the tenth day on. There are two days when distinct losses are registered in the one period group, and these would be very undesirable in any learning because of the bad effect on the attitude of the learner toward his work when he fails to make progress.

The experiment is useful in a number of other ways. Plotting of the individual records, for instance, gives excellent illustrative material for the study of individual differences. In the usual class there will probably be found curves which will illustrate all the common characteristics of practice curves such as plateaus, daily fluctuations, and so on.

If the subjects keep careful records of the conditions under which they work at different times during the practice period some interesting facts may be obtained concerning the effect of various factors on the performances of individuals. Changes resulting from interest, fatigue, distractions, time of day when practice is done, and like circumstances are likely to appear, and add greatly to the value and interest of the work.

The scores in the various tests furnish excellent material for illustration of the mathematical problems of measurement. The

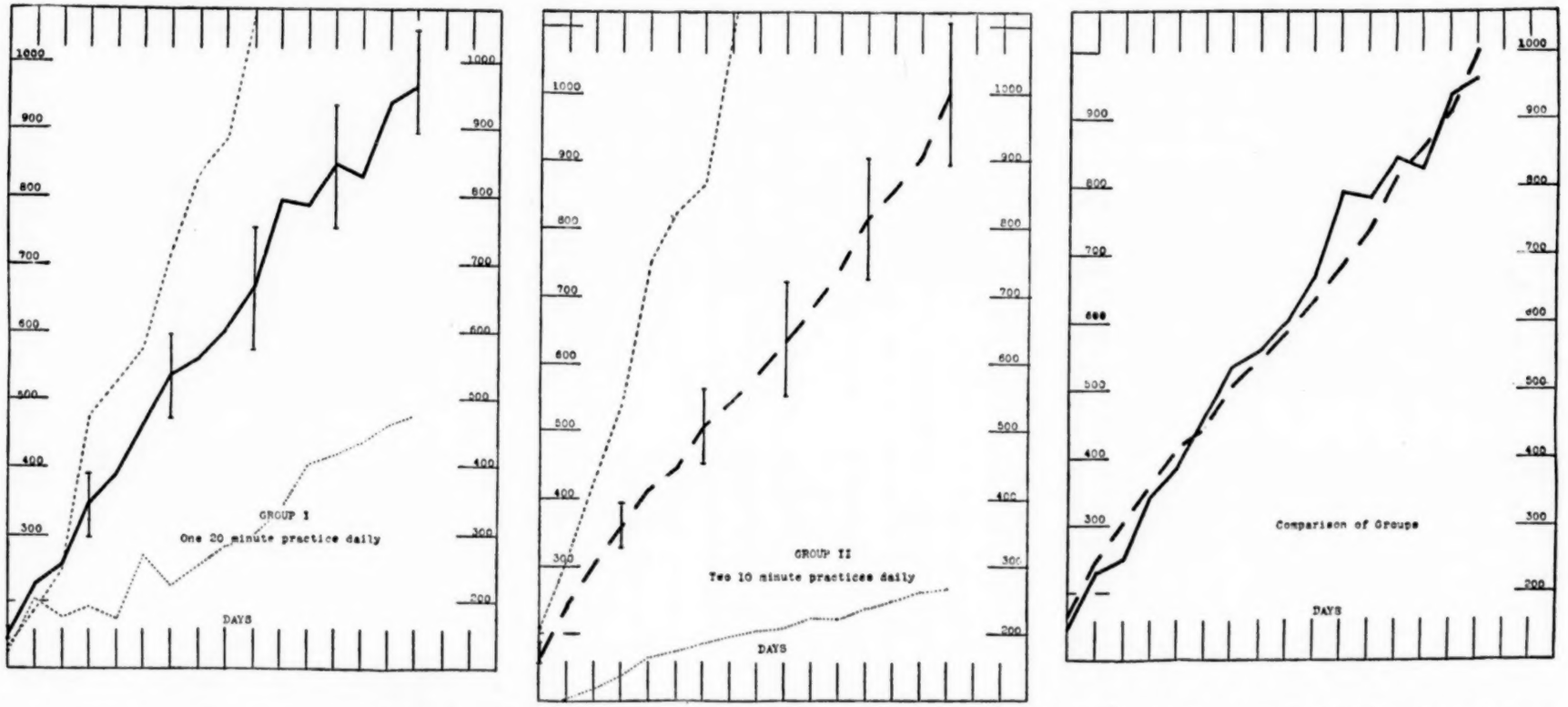

PLATE I.—Daily averages of Group I, with best and poorest records of the group. The perpendicular lines every third day show the average deviations on those days.

PLATE II.—Daily averages of Group II, with best and poorest records of the group. The perpendicular lines every third day show the average deviations on those days.

PLATE III.—Comparison of the two groups to determine the better time distribution in practice periods.

data obtained proved especially useful in illustrating to the class the methods of numerical and graphic correlation.

While it is believed that the findings will ordinarily be in agreement with the more general and more extensive experiments in the same field, it is possible that there will be differences in some of the select groups which make up college and normal school classes. Care should be taken, therefore, to make it clear, when this or a similar experiment is used, that the results do not prove or disprove the various educational theories concerning which they furnish evidence.[1]

---

[1] This experiment was originally begun as a contribution to the work of a committee of the American Psychological Association on Class Experiments in Psychology. If there is sufficient demand, the materials will be printed in quantities and supplied at cost to those interested.

LANGUAGE ERROR TESTS

G. M. WILSON

Boston University

For 10 years the writer has been interested in language errors and for the last 5 years has been interested actively in a test that would show the ability of grade and high school pupils to recognize and correct the most common errors of the English language. The purpose of this article is to describe briefly an attempt to develop such language error tests and to show as final results three simple tests that are practically equal in value.

Determining the Errors.—A language test to be of any great value must be based squarely upon the most common mistakes made by school pupils. This has not always been done in attempts to make language error tests; partly, because data were not available until recently. Now, however, the most common errors have been charted so fully that it is no longer a matter of opinion but a matter of record.[1]

The errors selected for the present test were chosen on the basis of the language errors studies at Connersville, Indiana; Kansas City, Missouri; Boise, Idaho; Iowa Consolidated Schools; and Cincinnati, Ohio.

Determining the Method of Testing.—In determining the method of testing, it was the writer's desire to place the language errors before the child in the form in which the children themselves make the errors. It is this situation of which the child needs to become conscious. This purpose has been accomplished by putting the tests in the form of ordinary compositions, as they might be written by children. It becomes the duty of the child to recognize the errors and to correct them. The child thus becomes his own teacher. It was soon found that the test appealed to children and that they responded to it eagerly.

Some may object that it is unwise to place incorrect language forms before children; that it is false pedagogy. It is evident, however, that the tests provide a good teaching situation since improvement as a result of giving the tests is exceedingly rapid. If the tests do bring the children up rapidly, making them conscious of what is necessary

---

[1] Wilson, G. M.: Locating the Language Errors of School Children, *Elementary School Journal*, Vol. XXI, December, 1920, pp. 290–296. This is a summary of several studies.

and creating the desired interest in correct language forms, then, certainly the tests are acceptable from a pedagogical standpoint.

Getting a Good Test.—The author, working with an advanced class in Education, listed the various language errors and especially noted the 30 or 40 that were most common. As a class exercise the attempt was then made to write compositions in each of which there should be about 30 of the most common language errors. Some of these compositions were worthwhile as a basis for the formulation of tests. Finally the author was able to put three of these compositions in form for use as tests. The three compositions were entitled, "Playing Marbles," "Strawberry Time," and "A Thanksgiving Dinner." The next step was to test these tests by using them upon public school pupils. The result of this test was to demonstrate clearly that there was a wide divergence in the grade-score values from the three tests. "Playing Marbles" showed by far the most desirable distributions. The preliminary testing of the tests gave the following tentative standards:

TABLE I.—TENTATIVE STANDARDS

Grade	Test 1 "Playing Marbles"	Test 2 "Strawberry Time"	Test 3 "A Thanksgiving Dinner"
III	6	8	10
IV	10	13	17
V	14	23	19
VI	15	24	21
VII	17	24	—
VIII	18	25	—

These standards were based upon the testing of a relatively small number of pupils—410 pupils for "Playing Marbles;" 269 pupils for "Strawberry Time;" and 452 pupils for "A Thanksgiving Dinner." It was evident, however, that "Strawberry Time" was too easy for upper grade pupils. With 26 corrections possible Grade V made a median score of 23; Grade VI, 24; Grade VII, 24; and Grade VIII, 25.

In view of the fact that "Strawberry Time" and "A Thanksgiving Dinner" are not acceptable as tests because they are too easy or do not show good distributions or proper slopes from one grade to another, it will be unnecessary to reproduce them here. "Playing Marbles," however, is so acceptable from every standpoint that a copy of it will

be of interest. The form, including directions as actually used, follows herewith:

Test 1.—Correcting Language Errors. (*A game*)

Name..... Grade.....
Town..... School.....
Date..... Age.....

Directions for the Game.—(To be read by the teacher, the pupils following.) This is a little *game* in which the pupil plays teacher, and corrects a composition written by a pupil. Correct by drawing a single line through words or expressions used incorrectly, and placing the correct words above them. For example if you had the following sentence to correct: "He has went home," you would correct it by drawing a single line through *went* and writing *gone* above it. Make all changes necessary to secure correctness. Work at your usual rate. You will be given reasonable time in which to complete your work. When you have finished, turn the sheet right side down and leave it on your desk. All will be permitted to finish the work unless too slow.

The composition which you are to correct follows herewith:

Playing Marbles

Marbles is a good game. I seen some boys playing the game yesterday. I went home to look for my supply of marbles. I couldn't find none, so I saw my father. I said to him: "Father, I ain't got no marbles. Will you give me a dime?" Father seen that I was in earnest, so he give me a dime. He done it willingly. Me and father is very good friends.

I started down the street. I had not went very far, when I met John Taylor. John he is a good friend of mine. He seen me leave my home, and had came to meet me. I owed him a dime, but he did not ask me to pay up. I guess he wanted me to have some marbles so as I could play with him. He had some marbles hisself.

I ask him to go to the store with me. "No," he replied, "I have got an errand to run. Can I play with you when I have did the errand?" We agreed and spent the entire afternoon together. We had lots of fun.

If the language test were to be given once only, it would be recommended that "Playing Marbles" be used as the test. The returns on it show very acceptable distributions, and good progress from one grade to another. It is a good measure of a child's ability to detect and correct errors in written language. It is so simple that it can be

used as low as Grade III; so difficult that few college students make perfect scores. Twenty-four is a perfect score.

Table II, which follows, shows a typical table of distribution for Test I, "Playing Marbles." There is the usual overlapping of ability from grade to grade but, nevertheless, good progress and no perfect scores. The returns clearly indicate a good test.

TABLE II.—TYPICAL DISTRIBUTION FOR SCORE OF RIGHTS
Test 1, "Playing Marbles."
(Duluth, Minn., 1918)

Score	Grades						
	III	IV	V	VI	VII	VIII	XI
0	3	3	1				
1	7	3					
2	7	2	1				
3	7	1	..	1			
4	7	6	1				
5	2	2	1				
6	3	3	3				
7	5	4	2				
8	1	5	2	5			
9	2	3	5	7	..	1	
10	5	7	6	4			
11	..	7	8	5	..	2	
12	1	3	8	11	1	2	
13	..	4	7	6	1	2	
14	..	5	11	6	..	1	2
15	..	..	13	10	2	6	2
16	..	2	10	16	2	8	
17	..	1	9	7	1	6	3
18	..	..	13	9	1	11	8
19	..	2	3	9	1	13	15
20	..	2	4	6	..	5	19
21	..	..	1	3	1	8	15
22	..	..	..	1	..	..	8
23	..	..	..	1	..	1	1
24							
Totals.....	50	65	109	107	10	66	73
Medians.....	4	10	14	15	16	18	20

Getting Three Tests of Equal Value.—It was at this stage in the development of the Language Error test that Dr. E. L. Thorndike of Teachers College was consulted. He commended the form of the test, its educational possibilities, and also its measurement possibilities. However, he asked that the research be continued until there should be three tests of equal value. This task proved a tedious one, involving a great deal of detailed statistical procedure and occasioning much delay. However, the final results gave three tests of practically equal value.

It will be of interest to note the procedure in securing these three equal tests. The first step was to evaluate all of the errors in the three previously used stories, "Playing Marbles," "Strawberry Time" and "A Thanksgiving Dinner." On the basis of tests that had been given to 1131 pupils in Sioux City, Iowa; Ambler, Pennsylvania; and Duluth, Minnesota, the error value of each error was figured for each grade. This gave for "Playing Marbles" a return as shown in Table III.

This table means that 83 per cent of the III*B* pupils failed to correct the first error in Test I, "Playing Marbles," *i.e.*, they failed to cancel "seen" in the first line and write "saw" above it. In the VI*A* grade, only 8 per cent of the pupils failed to correct this error. The errors differ greatly in difficulty, as shown by Table III. Errors 1, 3, 5, 11, 12, and 19 show gradual reduction in the higher grades. Error 17, on the other hand, is not recognized by pupils in Grade VI and below, while errors 22 and 24 are seldom recognized. It was to ascertain this difference in error value that Table III on "Playing Marbles" and similar tables on Tests 2 and 3 were constructed.

With the value of each error thus figured, it was possible to re-arrange the errors so as to distribute them equally for new stories. This re-arrangement with values for the different grades and total values is shown herewith in Tables IV, V, and VI. The initials in the following tables are explained as follows: P. M.—"Playing Marbles;" S. T.—"Strawberry Time;" T. D.—"Thanksgiving Dinner." The two errors added after the first total in each case were for the purpose of making the error values of the different stories more nearly equal and particularly to help the slope. The resulting totals are surprisingly close together and the slope from grade to grade is surprisingly uniform:

TABLE III.—PER CENT OF ERROR
Test 1. "Playing Marbles"
(Duluth)

Error	Grade								
	III*B*	III*A*	IV*B*	IV*A*	V*B*	V*A*	VI*B*	VI*A*	Total
1. *Seen* for *saw*	83	68	34	22	32	0	12	8	259
2. *None* for *any*	83	66	68	37	18	22	25	13	332
3. *Ain't* for *have*	83	62	28	11	18	11	12	8	233
4. *Got*	83	68	66	67	63	71	77	29	524
5. *No* (double neg.)	100	79	50	37	39	26	40	8	379
6. *Seen* for *saw*	100	70	65	30	46	26	23	21	381
7. *Give* for *gave*	100	95	59	44	53	39	28	25	443
8. *Done* for *did*	100	90	90	55	57	61	58	46	557
9. *Me* for *I*	100	77	59	55	53	50	33	21	448
10. *Me and father*	100	84	78	88	57	50	44	21	522
11. *Is* for *are*	100	81	59	38	46	33	19	8	384
12. *Went* for *gone*	83	77	50	32	25	33	26	8	334
13. *John, he*	100	77	97	55	58	61	65	13	526
14. *Seen* for *saw*	83	70	56	30	39	17	12	17	324
15. *Came* for *come*	83	81	66	44	53	50	42	17	436
16. *Pay up* for *pay*	100	77	81	67	61	88	44	38	556
17. *Guess*	100	100	100	100	100	100	100	100	800
18. *As* (superfluous)	100	77	87	67	61	77	56	42	562
19. *Hisself* for *himself*	100	92	62	37	32	39	28	4	394
20. *Ask* for *asked*	100	92	99	74	61	50	44	34	554
21. *Have got* for *have*	100	86	99	93	93	88	93	55	707
22. *Can* for *may*	100	92	100	100	100	94	96	76	758
23. *Did* for *done*	100	86	74	52	46	39	23	13	433
24. *Lots of*	100	95	100	93	86	100	89	55	718

TABLE IV.—STORY NO. A. BASIS FOR IN RECLASSIFIED ERRORS

Source	Value	III*B*	III*A*	IV*B*	IV*A*	V*B*	V*A* ±	VI*B* ±	VI*A* ±	VII*B* ±
S. T. 2.....	18	11	3	0	4	0	0	0	0	0
T. D. 23.....	189	59	44	15	18	17	21	9	6	
P. M. 3.....	233	83	62	28	11	18	11	12	8	
S. T. 4.....	217	66	55	34	44	4	0	14	0	0
T. D. 1.....	233	43	64	35	12	19	33	9	18	
T. D. 18.....	238	52	45	24	15	32	29	23	18	
S. T. 8.....	256	77	76	34	40	4	8	10	0	7
S. T. 17.....	269	77	59	52	68	8	0	5	0	0
S. T. 13.....	308	89	76	41	64	8	4	14	12	0
T. D. 7.....	446	76	62	52	45	63	50	50	48	
T. D. 3.....	370	67	69	52	24	47	44	38	29	
T. D. 21.....	379	59	62	46	52	49	42	25	44	
P. M. 6.....	381	100	70	65	30	46	26	23	21	
S. T. 20.....	406	89	86	57	64	19	6	33	53	0
P. M. 23.....	433	100	86	74	52	46	39	23	13	
T. D. 26.....	449	63	71	49	58	52	60	52	44	
T. D. 17.....	481	94	69	68	61	48	54	39	48	
S. T. 18.....	472	91	93	72	84	31	4	50	47	0
P. M. 13.....	526	100	77	97	55	58	61	65	13	
P. M. 4.....	524	83	68	66	67	63	71	77	29	
P. M. 18.....	562	100	77	82	67	61	77	56	42	
P. M. 10.....	522	100	84	78	88	57	50	44	21	
T. D. 22.....	633	85	91	79	82	81	73	68	74	
T. D. 11.....	669	97	80	71	82	91	94	77	77	
S. T. 11.....	645	82	83	92	88	81	67	55	47	50
S. T. 5.....	317	75	76	56	64	0	17	17	12	0
Total.....	10,176	2018	1788	1419	1339	1003	941	888	724	57
P. M. 7.....	443	100	95	59	44	53	39	28	25	
S. T. 19.....	227	75	62	23	44	2	4	5	12	0
Total.....	10,846	2193	1945	1501	1427	1058	984	921	761	57

TABLE V.—STORY NO. B. BASIS FOR IN RECLASSIFIED ERRORS

Source	Value	III*B*	III*A*	IV*B*	IV*A*	V*B*	VA ±	VI*B* ±	VI*A* ±	VII*B* ±
S. T. 22.....	164	62	41	10	24	19	6	2	0	0
S. T. 6.....	167	44	41	26	24	15	8	2	0	7
T. D. 13.....	199	37	47	28	24	16	21	16	10	
S. T. 1.....	197	57	55	23	52	0	0	10	0	0
S. T. 12.....	237	80	66	18	48	8	4	6	7	0
T. D. 10.....	250	56	58	26	33	35	17	9	16	
P. M. 1.....	259	83	68	34	22	32	0	12	8	
T. D. 9.....	312	69	67	41	36	28	29	18	24	
S. T. 26.....	321	82	79	46	56	15	0	19	24	0
P. M. 2.....	332	83	66	68	37	18	22	25	13	
S. T. 24.....	372	84	76	46	76	23	0	36	24	7
T. D. 16.....	387	63	69	45	45	48	48	30	39	
P. M. 5.....	379	100	79	50	37	39	26	40	8	
P. M. 19.....	394	100	92	62	37	32	39	28	4	
S. T. 7.....	438	88	100	57	92	0	12	36	53	0
P. M. 11.....	384	100	81	59	38	46	33	19	8	
T. D. 30.....	443	76	64	61	45	53	56	43	45	
P. M. 15.....	436	83	81	66	44	53	50	42	17	
T. D. 20.....	455	80	75	54	48	57	60	41	40	
S. T. 23.....	484	91	89	66	96	35	21	55	24	7
T. D. 25.....	506	89	82	54	48	65	50	66	52	
T. D. 12.....	548	89	62	66	70	72	73	55	61	
P. M. 20.....	554	100	92	99	74	61	50	44	34	
T. D. 15.....	573	91	96	80	61	61	63	55	66	
T. D. 29.....	691	94	85	79	76	92	89	86	90	
P. M. 21.....	707	100	86	99	93	93	88	93	55	
Total.....	10,189	2081	1897	1363	1336	1016	865	888	722	21
P. M. 6.....	381	100	70	65	30	46	26	23	21	
S. T. 13.....	308	89	76	41	64	8	4	14	12	0
Total.....	10,878	2270	2043	1469	1430	1070	895	925	755	21

TABLE VI.—STORY NO. C. BASIS FOR IN RECLASSIFIED ERRORS

Source	Value	III*B*	III*A*	IV*B*	IV*A*	V*B*	V*A* ±	VI*B* ±	VI*A* ±	VII*B* ±
S. T. 15.....	124	39	31	15	24	8	0	0	0	7
S. T. 3.....	169	54	24	25	36	8	4	12	6	0
S. T. 10.....	193	55	48	31	48	0	0	5	6	0
S. T. 14.....	138	39	24	67	8	0	0	0	0	0
S. T. 19.....	227	75	62	23	44	2	4	5	12	0
T. D. 2.....	268	56	69	38	27	21	35	9	13	
T. D. 14.....	269	46	64	37	27	32	31	14	18	
T. D. 5.....	324	74	58	34	33	32	40	27	26	
T. D. 6.....	325	65	58	35	48	36	33	21	29	
P. M. 12.....	334	83	77	50	32	25	33	26	8	
T. D. 28.....	341	69	65	42	36	49	33	27	20	
T. D. 19.....	370	69	62	42	45	43	42	30	37	
S. T. 25.....	423	88	86	62	96	31	16	26	18	0
S. T. 9.....	423	80	100	52	96	15	16	33	24	7
T. D. 31.....	403	81	80	59	30	47	42	30	34	
T. D. 24.....	424	89	76	49	64	48	44	23	31	
P. M. 9.....	448	100	77	59	55	53	50	33	21	
P. M. 7.....	443	100	95	59	44	53	39	28	25	
S. T. 21.....	503	88	100	59	88	26	12	48	82	0
S. T. 16.....	526	86	93	75	96	8	33	40	88	7
T. D. 4.....	537	76	75	65	73	75	67	43	63	
T. D. 27.....	522	87	78	68	61	63	65	55	45	
P. M. 16.....	556	100	77	81	67	61	88	44	38	
P. M. 8.....	557	100	90	90	55	57	61	58	46	
P. M. 24.....	718	100	95	100	93	86	100	89	55	
P. M. 22.....	758	100	92	100	100	100	94	96	76	
Total.....	10,323	1999	1856	1417	1426	979	982	822	821	21
P. M. 5.....	379	100	79	50	37	39	26	40	8	
S. T. 22.....	164	62	41	10	24	19	6	2	0	0
Total.....	10,866	2161	1976	1477	1487	1037	1014	864	829	21

SOME EVIDENCE OF AN ADOLESCENT INCREASE IN THE RATE OF MENTAL GROWTH

KATHERINE MURDOCK

Punahou School, Honolulu, Hawaii

AND

LOUIS R. SULLIVAN

American Museum of Natural History, New York

Many studies of mental growth have appeared very recently in our journals of psychology and education. Various phases of the subject have been attacked and defended by psychologists of note. The major issues of dispute seem to have centered about: first, the constancy of the ratio of mental development to age; that is, the constancy of the IQ, over various intervals of time and for various degrees of mental attainment; second, the comparative variability in mental development of the two sexes and of children of different life ages; and third, the limit of mental development, for normal and subnormal subjects. Such an aroused interest in the theoretical, as well as the practical, problems of mental growth will undoubtedly result in new generalizations concerning the whole topic, and the special phase of the subject with which we are here dealing will again come into prominence.

That the problem of the increase in rate of development at adolescence should for some time have been so neglected seems perhaps strange to the student of the physical development of the body. However, obvious as it is that study of adolescent changes in rate of growth is one of the primary tasks of any development study, there are good reasons why psychologists have been side-stepping that issue somewhat.

Probably one reason for the popular neglect of our subject is that mental units, sufficiently accurate for measuring such fine distinctions as our subject calls for have not yet been found. Mental development is at present usually measured in terms of mental age. That mental age steps are equal in amount is a fact not theoretically insisted upon by Terman,[1] nor practically believed in by Kelley.[2]

[1] Terman, L. M.: *Journal of Educational Psychyolog*, September, 1921. "The Binet type of scale does not necessarily presuppose equality of mental steps."

[2] Kelley, T. L.: *Journal of Educational Research*, p. 239, October, 1921. "I would say with reference to scales of the Binet type which assume equivalence of successive age intervals, I think we already have abundance of evidence to refute assumption."

In order to show changes in rate of development accurately, it is, of course, necessary that the measuring scale used *shall* be made up of equal units. Objectivity and self-equality of units, such as are illustrated by the inch and the pound, will probably never be attained in such full measure in mental measuring rods, but the earnest hope and belief of all psychologists is that mental measures may, in the future, more closely approach these ideals. Some writers have suggested that certain forms of "point scales" already excel the Binet mental age method in providing equal measuring units. Full confidence in such measures must wait upon the settlement of controversies such as have been engaged in by Freeman and Peterson, as to the proper method of equating "time" and "work." Possibly some form of the method suggested by Thurstone, the determination of standards for each mental age that state the percentage of unselected children at each life age who reach or exceed that grade of intelligence, would help to solve the equal unit problem. At any rate, it is not solved at present. This fact offers good justification to psychologists for having, in the main, avoided attempts at measuring adolescent growth accelerations. The direct use of the IQ is, of course, out of the question, since this is a measure of relative brightness, not of mental status.

Another, and probably more important, reason, why psychologists have not paid much attention to changes in rate of mental growth at adolescence, is a very practical one. The science of mental measurement is young. Matters concerning really large differences in mental ability are still unknown to many laymen. Measurers of mentality have been busy pointing these out and doing practical, necessary service. Early in the short period during which the IQ has been studied, it was conclusively shown by Terman that no *large* adolescent spurts exist. Since that was done in 1916, until perhaps within the last 12 months, little time has been spared by the psychologists studying mental differences from the practical duty of dealing with large differences in mentality, to the mere theoretical pursuit of measuring small fluctuations in the rate of mental growth.

Now the time seems to have arrived when mental growth curves should be more minutely studied. Until better measuring units are available, various devices will have to be used to show up certain features of these curves.

To study the adolescent growth curve, we have employed the simple device of making comparison between the physical and mental sex differences which occur during the development ages. Boas, in an

article on the Growth of Children, *Science*, Vol. XXXVI, has set forth clearly and concisely the laws of physical growth. The curve for rate of growth, in almost all, if not all, organs and parts of the body, he shows to have two modes. The higher mode occurs during fetal life; the second, and lesser, occurs shortly before sexual maturity is reached. Since adolescence occurs at different ages in the two sexes, the pre-adolescent increase in rate of growth occurs at different ages—that in girls occurring about 2 years earlier than the corresponding acceleration in boys. The age for minimum increase in annual growth is given by Boas as 10.3 for boys and 8.2 for girls. The age for maximum increase is 13.2 for boys and 11.2 for girls. With these facts about sex differences in physical traits in mind, it occurred to us that if a similar sex difference were found to exist in mental ability, girls exceeding boys mentally during the same ages in which they exceed them physically, that this would constitute some evidence that an adolescent increase in rate of growth is a feature of mental as well as of physical development. We already had at hand measures used by us for another study, which made this comparison between physical and mental age-sex differences very easy to make. Records of 580 boys and girls, who ranged in age from 6 to 18, were used. All of these subjects at the time the measures were made were pupils of Punahou School, Honolulu, Hawaii. This is a private school which carries the pupils from the first grade through high school. All of the 580 pupils were American or British children, of Northern European descent. Most of them had lived all their lives in the Hawaiian Islands. The measures which we had at hand were made originally by us separately for independent purposes. The anthropological data were collected by Doctor Sullivan, the mental measures by Miss Murdock. The latter measurements were made by the use of group tests. The Otis Primary Test was used for Grades I to III; the National Intelligence Tests, Forms A and B, for Grades III to IX; the Terman Group Test, for the four high school grades. Mental ages were assigned to individuals on the basis of norms furnished by the authors or publishers of the tests. Adjustment was made in order to bring ages derived from the three different tests all to the standard of the National Intelligence Tests. Ages above and below those for which norms were available, were estimated from assumptions concerning the normal curve of distribution. IQs were found for each subject by dividing the mental age by the life age. (It is true that some doubt has been thrown by Freeman and others upon the permissibility of using mental ages,

derived from group tests, to obtain IQs. For the purposes for which we have used these IQs, however, we believe there can be no objection to the way in which we have derived them.) In the case of pupils from Grades III and IX, who received mental age ratings from two tests, an average of the two was used. The measures for weight were expressed in terms of pounds; those for stature, in terms of centimeters.

The results of the comparison of physical and mental measures for the two sexes at successive ages, are given in the accompanying tables. The average weight, stature, and intelligence quotient for boys and for girls at each age from 6 to 18 are given, and also the amount in each age group by which the girls exceed the boys, or the boys the girls, in weight, stature, and IQ. In the case of the IQ, the excess of girls over boys, and of boys over girls, is given for each age separately, and also as smoothed averages, in which the average given for each age group is in reality the average for the boys (or the girls) of that age group in combination with the age just younger and the one just older. Comparison of the excess columns is very striking in revealing that there is a similarity between the mental and physical ages of development. If we confine our attention to the "smoothed" averages for the IQs, and compare these with the physical measures, we find that the direction of excess, of boys over girls, or vice versa, at different ages, is as constant for mental and physical measures as it is between the two physical measures themselves. From 8 years of age until 13, the girls excel, physically and mentally. Thereafter they are behind the boys. The rough, or unsmoothed, excesses tell about the same story. The greatest exception occurs at the age of 18, where girls are seen to excel the boys, mentally, by an average of 2.4 IQ. It occurs to us as possible that the brighter of the 18-year-old boys may be sent to college more often than the bright girls of 18—parents disliking to have their daughters go so far from home at this early age. However this is only a supposition. The smallness of the groups naturally would result in irregular results for differences which are so small as those which we are attempting to measure. We would not, in fact, feel justified in presenting our data at all, based, as they are upon such limited numbers of cases, were it not for the fact that so many age groups unite in confirming the evidence.

Our own interpretation of our results is that they furnish evidence that a pre-adolescent increase in rate of mental growth occurs at the same time in the development of each sex that the physical increase in

development occurs. Whether or not boys excel girls mentally, as they do physically, after, and to a slight extent, before the adolescent growth periods, is a question on which our data hardly shed enough light for us to form an opinion. Neither do we feel justified in assuming anything about the amount of the adolescent mental growth acceleration, except that it is probably much smaller, comparatively, than the physical "spurt." The unit of mental age scales, by its very definition, is of such a nature that it tends to conceal any differences in rate of mental growth. Eleven years mental age means the mental age of the average 11-year-old child. If, on the average, children should develop little mentally from 10 to 11 years of age, and develop much from 11 to 12, properly arranged mental scales, of the age standard type, would entirely conceal such change in rate of development. (The IQ as a measure, is similarly limited, with the additional restriction, when it is used for purposes of studying mental development, that only subjects of the same life age can be compared. In our present study, we have used IQs, instead of mental age, only because our measures already had been converted into this form for other purposes. Since our comparisons all are between groups composed of individuals, who are of the same life age, the results are identical, whether mental age or IQs are used.) Since then changes in mental growth are concealed, rather than shown up, by the use of mental scales, it is impossible for us to arrive at a decision as to the amount of adolescent growth acceleration. Its existence at all, by the use of such scales, could not have been determined, were it not for the sex differences which we found. Mental scales have been standardized by the use of results obtained from the scores of both boys and girls.

On the whole our results seem to be in harmony with those of other investigators. Porteus, in his maze studies, found some correlation between physical and mental development, with a consequent superiority, in mental ability, of girls to boys from the age of 11½ to 13 years, and inferiority at both earlier and later ages. Yerkes and Bridges, by the use of their Point Scale for Measuring Intelligence, found boys superior to girls from 8 to 11, girls superior at 12 and boys again superior from 13 to 15. That the ages for the girls' superiority in our study come somewhat earlier than in these others, is possibly due to the fact of the early physical development, brought about by the warm climate or the social status of our subjects. Terman's findings seem also to yield substantially the same results as ours On p. 75 of "The Stanford Revision of the Binet-Simon Scale" he says:

Age	Averages: Number of cases		Weight in pounds		Stature in centimeters		IQ		Sex differences in averages: Excess in weight		Excess in stature		Excess in IQ		Excess in IQ (smoothed average)	
	Boys	Girls	Boys	Girls	Boys	Girls	Boys	Girls	Girls exceed boys	Boys exceed girls	Girls exceed boys	Boys exceed girls	Girls exceed boys	Boys exceed girls	Girls exceed boys	Boys exceed girls
6	8	12	50	48	121	119	109.1	104.9	..	2	..	2	....	4.2	...	
7	22	18	54	51	125	122	112.7	107.7	..	3	..	3	....	5.0	...	2.2
8	18	25	58	59	130	129	99.7	103.2	1	..	..	1	3.5	...	2.0	
9	29	23	63	68	133	136	102.8	110.1	5	..	3	..	7.3	...	6.3	
10	32	22	68	75	136	142	105.0	115.6	7	..	6	..	10.6	...	5.7	
11	33	17	81	89	145	149	113.1	113.1	8	..	4	..	0.0	...	4.9	
12	26	28	87	92	151	152	109.0	113.2	5	..	1	..	4.2	...	1.2	
13	27	28	97	111	155	160	113.2	113.0	14	..	5	..	...	0.2	1.0	
14	34	35	112	113	164	161	110.0	109.6	1	..	..	3	....	0.4	...	2.0
15	24	19	129	118	171	164	110.9	106.0	..	11	..	7	....	4.9	...	1.7
16	18	32	135	123	173	164	107.9	108.7	..	12	..	9	0.8	...	...	2.6
17	14	18	135	122	175	165	114.4	108.5	..	13	..	10	....	2.9	...	0.1
18	7	13	140	122	175	165	109.9	112.3	..	18	..	10	2.4			

> "In the main, therefore, the school progress of our subjects agrees with the intelligence tests, with the teachers' estimates of intelligence, and with the teachers' judgments of the quality of the school work, in showing a sex difference which is in favor of the girls before 14, and in favor of the boys thereafter."

This statement, however, is immediately followed by Terman with an explanation of what he thinks is the probable reason for his findings. He believes the apparent superiority of boys over 13, to girls, to be due to the effect of selection of his subjects. All of these were pupils in the elementary school, and his belief is that more girls than boys, of 14 years of age, had been advanced to high school. His final conclusion therefore is different from ours. It is that "the only possibility seems to be that the apparent superiority of boys at the age of 14, as well as also their diminished inferiority at 13, is due solely to the uneven selection which has taken place at these ages." However this may be in the case of Terman's subjects, it certainly is not true for ours that the superiority of the boys after 13 years of age is due to a selective influence, which places more girls in high school, for in our study high school students as well as those in elementary school are tested. All pupils of the ages 6 to 18 in the whole school, which consists of 12 grades, were included in our study (except those of other races). Another important study, whose results harmonize with our conclusions, is Mrs. Pressey's study of sex differences, in which she finds girls slightly superior to boys in mental ability. Mrs. Pressey's pupils were elementary school pupils, therefore mostly below 14 years of age.

We hope that in the future workers who compare the mental ability of the two sexes, will present their results in such a way that the sexes can be compared age for age during the developmental years. Further studies of this sort, including measures of many more subjects, alone can corroborate, or refute, our tentative conclusion that for both sexes there is an increased rate of mental as well as physical development for several years prior to the attainment of sexual maturity.

TENTATIVE ORDER OF DIFFICULTY OF THE TERMAN VOCABULARY WITH VERY YOUNG CHILDREN

MARGARET V. COBB

Institute of Educational Research, Teachers College, Columbia University

In the course of a study of group intelligence tests in Grade I, which involved Binet examining, it became evident that the words in the Vocabulary Test (Stanford Revision) show an order of difficulty for small children which is noticeably different from that indicated on the blank. Tabulation of the scores on the separate words gave the following information, which may prove useful to others who are examining kindergarten and Grade I children.

The examining was done by five persons, including myself, the others being Miss Helen Davis, Director of Measurements and Special Education in the public schools of Jackson, Michigan, through whose courtesy the records were made available, and three of the kindergarten teachers of Jackson whom she had trained for the work and who had shown special aptitude for it. The children were examined in the kindergarten during the first semester of the school year 1920–1921, and in Grade I during the second semester of the same year. With a few children the test was begun with word 6 in each list; the previous words were omitted but given credit, if there were no failures before word 11 ("Roar" and "haste" were, however, always given.). The test was stopped when the child missed five successive words in each list. Three or four of the scoring conventions observed in addition to the rules in "The Measurement of Intelligence" should probably be noted here, since a different rule might have altered the result:

Bonfire.—Full credit, any definition which gave a hint of distinction between a bonfire and other fire, as "A fire outdoors," "A big fire," "Go burn things up," etc. Half credit, definitions of fire alone, as "Furnace fire," "Fire," "In a stove," etc.

Haste.—Full credit for definition involving hurry or speed. Half credit for quotation from school song ("haste away," etc.) or definition as "go," "fly," etc.

Afloat.—Full cedit for definition involving floating on surface. Half credit for definition as moving along, swimming, carried along, etc.

Eyelash.—Half credit for "hair over eye" if child points to eyebrow instead of lashes.

The distributions of total scores on the vocabulary test, and of chronological and mental ages at the time of examination, are shown

in Tables I, II and III. Table IV gives the data concerning each word of the vocabulary, including the number of times each was given and the number of successes, of failures, and of half credits in the kindergarten and the lower and upper Grade I. With the present word order, the following words, unknown to any of these children, were needlessly asked:

First Column	Times Asked	Second Column	Times Asked
16. skill	35	19. forfeit	7
17. ramble	18	20. sportive	5
18. civil	14	23. shrewd	1
21. juggler	7	24. repose	1
22. regard	4	25. peculiarity	1
23. stave	1		
24. brunette	1		
25. hysterics	1		

On the basis of these figures, it is suggested that when the vocabulary test is given to young children the following word-order will be found useful, and will give greater certainty in making inferences as to when it is safe to omit the easiest words, and when the child has been carried far enough down the list. It also makes the two lists more nearly equal in difficulty, so that the use of one alone is somewhat more accurate. (It is my belief that this procedure is, however, almost never advisable.) Mimeographed sheets may easily be made up in this order, or in whatever similar order anyone wishes to derive from the data.

	Credits		Credits
1. straw	104	1. orange	106
2. envelope	103	2. bonfire	97
3. gown	83	3. puddle	93
4. tap	87½	4. rule	81
5. scorch	73½	5. roar	73
6. eyelash	66	6. pork	55
7. afloat	39	7. health	45
8. impolite	32	8. plumbing	32
9. copper	23	9. haste	23
10. nerve	14	10. guitar	21
11. dungeon	9½	11. muzzle	13
12. curse	9	12. misuse	7
13. southern	7	13. snip	6
14. lecture	4	14. reception	4½
15. mellow	4	15. noticeable	3½
16. insure	3	16. quake	2½
17. outward	2½	17. treasury	2½
18. apish	2	18. crunch	1
19. ramble	0	19. majesty	

TABLE I.—DISTRIBUTION OF TOTAL SCORES, VOCABULARY

Total scores	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23
Kindergarten	1	..	3	..	4	6	..	..	1	1	1	..	1	..	..	..	1	..	..	..
IB	..	..	3	..	3	3	2	6	5	4	2	3	1	2	..	..	1	..	1	1
IA	..	1	..	..	5	2	4	7	7	1	5	3	5	3	3	..	1	1	2	..
Total	1	1	6	0	12	11	6	13	13	6	8	6	7	5	3	0	3	1	3	1

TABLE II.—DISTRIBUTION OF MENTAL AGES AT TIME OF EXAMINATION

Mental ages	5^{8}	5^{10}	6^{0}	6^{2}	6^{4}	6^{6}	6^{8}	6^{10}	7^{0}	7^{2}	7^{4}	7^{6}	7^{8}	7^{10}	8^{0}	8^{2}	8^{4}	8^{6}	8^{8}	8^{10}	9^{0}	9^{2}
Kindergarten	2	1	4	2	3	1	4	1	1													
IB	..	..	1	1	1	2	1	3	4	4	4	3	4	2	3	2	0	2				
IA	..	..	..	..	1	3	3	4	5	4	10	8	2	0	4	1	3	1	0	0	0	1
Total	2	1	5	3	5	6	8	8	10	8	14	11	6	2	7	3	3	3	0	0	0	1

TABLE III.—DISTRIBUTION OF CHRONOLOGICAL AGES AT TIME OF EXAMINATION

Chronological ages	5^{2}	5^{4}	5^{6}	5^{8}	5^{10}	6^{0}	6^{2}	6^{4}	6^{6}	6^{8}	6^{10}	7^{0}	7^{2}	7^{4}	7^{6}	7^{8}	7^{10}	8^{0}	8^{2}	8^{4}	8^{6}	8^{8}	8^{10}	9^{0}
Kindergarten	1	3	1	4	4	3	2	0	1															
IB	..	..	..	..	1	5	1	3	9	1	9	2	2	2	0	1	0	0	0	0	1			
IA	..	..	..	..	..	..	..	..	1	8	11	7	6	3	6	2	3	1	0	1	0	0	0	1
Total	1	3	1	4	5	8	3	3	11	9	20	9	8	5	6	3	3	1	0	1	1	0	0	1

TABLE IV

First Column

Word order on blank	Full credit			Half credit			No credit			Times asked			Total credits	Total times asked
	Kinder-garten	IB	IA	Kinder-garten	IB	IA	Kinder-garten	IB	IA	Kinder-garten	IB	IA		
1. gown	14	28	41	..	..	..	5	9	9	19	37	50	83	106
2. tap	12	29	46	..	1	..	7	7	4	19	37	50	87½	106
3. scorch	12	25	35	1	..	2	6	12	13	19	37	50	73½	106
4. puddle	18	32	43	..	..	..	1	5	7	19	37	50	93	106
5. envelope	18	36	49	..	..	..	1	1	1	19	37	50	103	106
6. rule	13	26	42	..	..	..	6	11	8	19	37	50	81	106
7. health	7	16	22	..	..	..	12	21	28	19	37	50	45	106
8. eye-lash	8	24	26	1	6	9	10	7	15	19	37	50	66	106
9. copper	4	8	8	2	1	3	10	28	39	16	37	50	23	103
10. curse	1	4	4	..	..	..	10	33	46	11	37	50	9	98
11. pork	0	10	20	1	22	27	0	5	3	1	37	50	55	88
12. outward	0	1	0	..	2	1	..	25	37	..	28	38	2½	66
13. southern	..	0	5	..	2	2	..	26	31	..	28	38	7	66
14. lecture	..	1	3	..	..	..	..	25	35	..	26	38	4	64
15. dungeon	..	3	6	..	1	..	..	22	33	..	26	39	9½	65
16. skill	..	0	0	..	..	..	..	16	19	..	16	19	0	35
17. ramble	..	0	0	..	..	..	..	7	11	..	7	11	0	18
18. civil	..	0	0	..	..	..	..	5	9	..	5	9	0	14
19. insure	..	1	1	..	1	1	..	12	23	..	14	25	3	39
20. nerve	..	0	3	..	9	13	..	5	11	..	14	27	14	41
21. juggler	..	0	0	..	..	..	..	1	6	..	1	6	0	7
22. regard	..	0	0	..	..	..	11	1	3	..	1	3	0	4

Second Column

Word order on blank	Full credit			Half credit			No credit			Times asked			Total credits	Total times asked
	Kindergarten	IB	IA	Kindergarten	IB	IA	Kindergarten	IB	IA	Kindergarten	IB	IA		
1. orange	19	37	50	..	..	..	..	..	..	19	37	50	106	106
2. bonfire	10	36	45	9	..	3	..	1	2	19	37	50	97	106
3. straw	18	37	49	..	..	..	1	..	1	19	37	50	104	106
4. roar	8	27	37	..	..	..	11	10	13	19	37	50	72	106
5. haste	1	1	15	1	1	10	17	35	25	19	37	50	23	106
6. afloat	2	6	15	7	11	14	10	20	21	19	37	50	39	106
7. guitar	4	5	12	..	..	..	15	32	38	19	37	50	21	106
8. mellow	2	1	1	..	..	..	14	36	48	16	37	49	4	102
9. impolite	3	13	15	..	2	..	11	21	34	14	36	49	32	99
10. plumbing	4	12	12	1	1	6	8	22	32	13	35	50	32	98
11. noticeable	0	1	1	..	1	2	1	22	36	1	24	39	3½	64
12. muzzle	0	4	9	..	..	..	1	22	36	1	26	45	13	72
13. quake	..	0	0	..	4	1	..	10	21	..	14	22	2½	36
14. reception	..	2	2	..	0	1	..	10	21	..	12	24	4½	36
15. majesty	..	0	0	..	0	1	..	9	19	..	9	20	½	29
16. treasury	..	0	1	..	1	2	..	7	16	..	8	19	2½	27
17. misuse	..	0	7	..	..	0	..	10	13	..	10	20	7	30
18. crunch	..	0	1	..	..	..	..	6	13	..	6	14	1	20
19. forfeit	..	0	0	..	..	..	..	1	6	..	1	6	0	7
20. sportive	..	0	0	..	..	..	..	1	4	..	1	4	0	5
21. apish	..	2	0	..	..	..	..	1	11	..	3	11	2	14
22. snip	..	4	2	..	..	..	..	3	15	..	7	17	6	24

The larger changes in order seem to be due to the fact that certain more abstract terms, such as *afloat, haste, mellow, noticeable, outward* and *skill*, are relatively more difficult at these early ages; and that some of the more concrete terms, such as *dungeon, envelope, eyelash, nerve, plumbing, pork* and *snip*, are learned earlier and are (relative to an older group having an equally small vocabulary) more likely to be known. Smaller changes may be due to the fact that the number of cases tabulated is small, though in general the lists from the subgroups confirm one another. The first four words appear in the same order with the kindergarten and I*B* grouped together as with the I*A* children; the first 13 words for the two groups are the same words, though the order is different. However, many tabulations are required before the best order for any limited group can be attained. It is suggested that others who have accumulated Stanford-Binet records for children of this age might tabulate and publish their data in such form that eventually they could be combined, and the best order finally determined.

SOME RETESTS WITH THE STANFORD-BINET SCALE

KATE GORDON

Bureau of Children's Aid, California Department of Finance

In a recent issue of the *Journal of Educational Psychology*, (September, 1921), an important summary was made of several studies on the constancy of the Stanford-Binet IQ as shown by retests. The number of cases reported in the present paper is small, but each additional bit of evidence on so vital a question seems to be worth having. Moreover, my results, as far as they go, indicate a difference in the results of retests on the two sexes.

All of the original tests on the group here reported, and all of the retests were made by me. The children tested are all of one race—the Hebrew—and all are living in the same environment, an orphanage for Jewish children. The original tests were made in October, 1918, when a complete mental survey of the institution was in progress. In December, 1919, those were examined who had been admitted since the previous date. In August, 1921, another complete survey was made of the same institution. The results given below include all of the children who were tested twice in the course of this procedure, namely, 44 persons. For 34 children the interval between tests was approximately 2 years 10 months; for 9 children about 1 year 8 months, and for 1 child, it was 1 year 3 months.

The number being small, I give the individual records in Table I.

In the cases of Boy No. 23, and Boy No. 27 I believe that the loss in IQ is due to the fact that these boys, who passed most of the 18-year-old tests of the scale, had no chance to earn a higher IQ because of the upper limit of the scale. On the other hand, Boy No. 17 also passed most of the 18-year-old tests, and had he been able to earn more, his gain in IQ would have been still greater.

The total distribution of changes in IQ is given in Table II. Their average is 6.8.

In spite of the occasional large changes indicated in Table II there is on the whole a very substantial agreement between the results of the first and second tests of this group, the correlation being r-0.84, as shown in Table III.

A glance at Table I reveals the fact that most of the losses were with the girls and most of the gains with the boys. I can assign no

TABLE I

Boy number	Age at Test I		Intelligence quotient on Test I	Age at Test II		Intelligence quotient on Test II	Points lost	Points gained
	Years	Months		Years	Months			
1	4	2	100	7	0	95	5	
2	5	0	110	7	11	112	..	2
3	5	8	100	7	4	91	9	
4	6	4	95	8	0	94	1	
5	7	3	89	8	11	93	..	4
6	8	5	85	10	1	98	..	13
7	8	5	72	11	2	69	3	
8	8	7	95	11	5	99	..	4
9	8	10	89	11	8	96	..	7
10	9	0	92	10	8	103	..	11
11	9	1	122	11	10	138	..	16
12	9	1	119	11	10	118	1	
13	9	9	95	12	7	87	8	
14	9	10	92	11	6	93	..	1
15	9	10	94	12	8	91	3	
16	10	4	100	13	2	99	1	
17	10	7	125	13	5	141+	..	16
18	10	7	94	13	5	106	..	12
19	11	5	100	14	3	110	..	10
20	11	5	91	14	2	105	..	14
21	11	6	107	14	4	108	..	1
22	11	7	71	12	10	77	..	6
23	11	10	130	14	7	122+	8	
24	12	0	79	14	10	83	..	4
25	12	7	98	15	5	111	..	13
26	12	7	98	15	5	101	..	3
27	12	11	126	15	10	117+	9	
28	13	0	92	14	8	99	..	7

Girl number	Age at Test I		Intelligence quotient on Test I	Age at Test II		Intelligence quotient on Test II	Points lost	Points gained
	Years	Months		Years	Months			
1	5	4	103	8	2	100	3	
2	6	4	76	7	11	80	..	4
3	6	4	79	7	11	78	1	
4	8	2	114	10	11	110	4	
5	8	3	109	11	0	106	3	
6	8	5	109	11	3	104	5	
7	8	11	121	11	9	118	3	
8	8	11	95	11	9	81	14	
9	10	7	83	13	5	75	8	
10	11	6	103	14	4	91	12	
11	11	8	89	14	6	80	9	
12	11	10	92	14	8	89	3	
13	12	10	101	15	8	98	3	
14	12	10	94	15	8	84	10	
15	13	2	116	16	0	93	23	
16	13	7	97	16	5	103	..	6

reason for this unless it be true that the girls come to an earlier stop in mental growth. Girl No. 15, who has a drop of 23 points seems an entirely normal child. Naturally one does not regard the matter as proved by 16 cases, but these are sufficiently arresting to suggest that it might be well for the records of boys and girls to be listed separately in reporting on the subject of retests.

"THE CONSTANCY OF THE IQ" AGAIN

FLORENCE M. TEAGARTEN

University of Pittsburg

All persons concerned with individual mental testing have been following with keen interest, in recent months, the discussion running through psychological literature relative to the constancy and diagnostic value of the IQ. Considerable evidence has been brought forth which claims to prove that the IQ can not be depended upon to give consistent results in subsequent tests. It appears, however, that in many of the investigations thus reported there have been present very conspicuous factors which have not only militated against the constancy of the IQ but also against the validity of any argument apparently inherent in the research.

One of these factors is the use of different tests or of different revisions of the same test. No very convincing scheme has been offered as yet for transmuting test results obtained by one revision or one test into supposedly comparable results by another revision or another test.

Again, some recent investigations conducted upon the feeble-minded claim to have proven that no dependence can be placed in the predictive value of the IQ. It is not the purpose of this report to dispute this point in the case of *institutional feeble-minded and of all ages*. For such cases the IQ may or may not be of significance but whether it be or not, are we justified in assuming that similar results are bound to accrue from tests of a school population; *e.g.*, since, it appears, we can not come to any agreement upon such fundamental questions as the nature of intelligence or the probable duration of its development it does seem a little beside the point to make sweeping assertions to the effect that laws of mental development of institutional feeble-minded as determined by their IQs are therefore the laws of mental development for the population at large.

The point which the writer here wishes to make is that what we probably need most right now is an *intelligent evaluation* of individual IQs before we can make intelligent comparisons of IQs. Probably sufficient evidence could be produced to show that in many cases where the bare IQ, stripped of all interpretation and standing in its mathematical nakedness, has appeared to change upon subsequent tests the very same IQ, clothed with a little common sense, could easily be identified later on. Two cases of this nature are here cited.

In November, 1921, two students taking training under the writer in administering Binet Tests (Standard Revision) examined, among others, two children in a certain practice school. The students had had several weeks of training. While the writer feels that there is always very grave danger in making serious comparisons of tests administered by different people, yet the dangers in this case were as small as it is ever possible for them to be since the students had their training from the one who made the retests and since the students' work was very carefully checked over with them on points of scoring, etc., following the examination. In March, 1922, for certain administrative reasons the writer was asked to go into the practice school and retest these two children, Jack and Mary.

The second examination showed that in the interval between November and March Jack's IQ had risen from 83 to 90+ and Mary's from 119+ to 128+. In the case of Jack the figures seemed to show that in something over 3 months he had risen from being a "dull normal" to a "normal" child and in exactly 4 months Mary had changed from a "superior" to a "very superior" child. Knowing the technique of the student testers, the writer did not feel warranted in passing off the matter by saying: "Oh, well, it is just a bit of the personal equation showing up." On the other hand, having a firm conviction in a reasonable constancy of the IQ she did not feel that the case could be dismissed without seeking some reason for this apparent lack of conformity.

In order that the reader may make the analysis for himself a brief chart of both tests for Jack and for Mary are here given. On the November test, Jack was chronologically 9-10 and mentally 8-2. On the March test his chronological age was 10-2 and his mental age 9-3, or a gain of 13 months. Mary, on the November test, was chronologically 7-7 and mentally 9-1. On the March test she was chronologically 7-10 and mentally 10-1, or a gain or exactly one year. The chart shows in the case of each part of the test the credit given (plus or minus) for that part together with the score for each portion of it. For example, in Year IX Test 2 (arrangement of weights) three trials are allowed. A plus score requires that two out of the three be correct. In the case of Jack's first examination —(− − −) means that he did not score on that test and that he failed all three trials. In his retest his score for the same test is +(+ − +) which means that his score was plus because he had correct two attempts out of three.

From a comparison of these reports for Jack it is evident that his

JACK

Test Nov. 22, 1921.
Chronological age 9-10: Mental age 8-2: Intelligence quotient 83.

Year VII

Tests all passed.

Year VIII

1. Ball and field	−	
2. 20-0	+	
3. Comprehension	− (−+−)	
4. Similarities two things	− (?−−−)	
5. Definitions	+ (++++)	
6. Vocabulary	+ (26)	

Year IX

1. Date	− (+++−)	
2. Weights	− (−−−)	
3. Makes change	+ (++−)	
4. Four digits backwards	− (−−−)	
5. Three words in sentence	+ (++−)	
6. Rhymes	+ (+++)	

Year X

1. Vocabulary	− (26)	
2. Absurdities	− (+−−++)	
3. Designs	+	
4. Reading and report	− (14 memories, 30″, 3 mistakes)	
5. Comprehension	− (+−−)	
6. Sixty words in three minutes	− (55)	

Years XI, XII

1. Vocabulary	− (26)	
2. Abstract words	− (+−−−?)	
3. Ball and field	−	
4. Dissected sentences	− (−−−)	
5. Fables	(Omitted)	
6. Five digits backwards	− (See IX, 4)	
7. Picture interpretation	− (?−− omitted)	
8. Similarities 3 things	− (See VIII 4)	

Years XIII, XIV

No tests given since all of XI, XII failed.

Test Mar. 6, 1922.
Chronological age 10-2: Mental age 9-3: Intelligence quotient 90+.

Year VII

Tests all passed.

Year VIII

1. Ball and field	+	Gain 2 months
2. 20-0	+	
3. Comprehension	+ (+++)	Gain 2 months
4. Similarities	− (+−−−)	
5. Definitions	+ (++++)	
6. Vocabulary	+ (27)	

Year IX

1. Date	− (++−+)	
2. Weights	+ (+−+)	Gain 2 months
3. Makes change	− (+−−)	Loss 2 months
4. Four digits backwards	+ (+−+)	Gain 2 months
5. Three words in sentence	+ (++ omitted)	
6. Rhymes	+ (+−+)	

Year X

1. Vocabulary	− (27)	
2. Absurdities	+ (++++ omitted)	Gain 2 months
3. Designs	+	
4. Reading and report	+ (14 memories, 25″, 0 mistakes)	Gain 2 months
5. Comprehension	− (−−−)	
6. Sixty words in three minutes	− (47)	

Years XI, XII

1. Vocabulary	− (27)	
2. Abstract words	− (+−−−?)	
3. Ball and field	−	
4. Dissected sentences)	− (−−−)	
5. Fables	− (+−−−−)	
6. Five digits backwards	− (−−−)	
7. Picture interpretation	+ (++++)	Gain 3 months
8. Similarities three things	− (−−− omitted)	

Total gain 13 *months*

Year XIII, XIV

Abbreviated test (4 tests) given—all failed.

MARY	
Test Nov. 7, 1921.	Test Mar. 7, 1922.
Chronological age 7-7. Mental age 9-1. Intelligence quotient 119+.	Chronological age 7-10. Mental age 10-1. Intelligence quotient 128+.
Year VIII	**Year VIII**
Tests all passed.	Tests all passed.
Year IX	**Year IX**
1. Date ..... − (+++−)	1. Date ..... + (++++) Gain 2 months
2. Weights ..... − (−+−)	2. Weights ..... + (−++) Gain 2 months
3. Makes change ..... + (++−)	3. Makes change ..... + (++−)
4. Four digits backwards ..... + (−++)	4. Four digits backwards ..... + (+−−)
5. Three words in sentence ..... − (+−−)	5. Three words in sentence ..... + (++−) Gain 2 months
6. Rhymes ..... + (+++)	6. Rhymes ..... + (+++)
Year X	**Year X**
1. Vocabulary ..... − (24)	1. Vocabulary ..... − (22)
2. Absurdities ..... + (++++++)	2. Absurdities ..... + (++−++)
3. Designs ..... +	3. Designs ..... +
4. Reading and report ..... − (14 memories, 40″, 0 mistakes)	4. Reading and report ..... + (10 memories, 23″, 0 mistakes) Gain 2 months
5. Comprehension ..... − (−+−)	5. Comprehension ..... + (++−) Gain 2 months
6. Sixty words in three minutes ..... − (55)	6. Sixty words in three minutes ..... + (60) Gain 2 months
Years XI, XII	**Years XI, XII**
1. Vocabulary ..... − (24)	1. Vocabulary ..... − (22)
2. Abstract words ..... − (−−−−−)	2. Abstract words ..... − (+−−−−)
3. Ball and field ..... −	3. Ball and field ..... + Gain 3 months
4. Dissected sentences ..... − (−− omitted)	4. Dissected sentences ..... − (−+−)
5. Fables ..... − (−−−−−)	5. Fables ..... − (−−−−−)
6. Five digits backwards ..... − (−−−)	6. Five digits backwards ..... − (−−−)
7. Picture interpretation ..... − (+−−−)	7. Picture interpretation ..... − (−−− omitted)
8. Similarities three things ..... + (−++−+)	8. Similarities three things ..... − (−+−−+) Loss 3 months
	Total gain 12 *months*
Years XIII, XIV	**Years XIII, XIV**
No tests given since all about 1 failed in Years XI, XII and nature of tests at this level indicated failure.	Abbreviated test (4 tests) given—all failed.

actual mental gain is much more nearly negligible than his 13 months accelerated credit would at first glance seem to indicate. In fact there are actually only two tests that were *entirely* failed the first time and *entirely* passed the second time, namely VIII-1, Ball and Field (counting 2 months credit) and XII-7 Picture Interpretation (counting 3 months credit). Five months credit then were undoubtedly gained in the interval of a little over 3 months. If Jack's intelligence were developing at anything like a normal rate we could probably admit a gain of 5 months in more than 3 months "and no questions asked." But where did his other 8 months credit come from? Ah! Here comes the need for analysis and interpretation, for some of them came from tests which he barely failed the first time (+ − −) and barely passed the second time (+ + −) and from similar "marginal" gains. For example, on the second test he received 2 months additional credit in the VIII Year Comprehension Test for a record of (+ + +) as against his previous record of (− + −). In fact, if his record on the second test had been only (+ + −) he would have received the additional 2 months credit since passing requires only 2 out of 3 successes at this point. Since, however, he had been able to pass even one of these tests in the first examination we can not say that he showed any unusual development to be able to pass two more 3 months later. Or again, consider Test IX-2 (Weights) where his first record was (− − −) and his second was (+ − +), and yet he gets 2 months additional credit for this partial ability which counts as a complete success. Exactly the same thing is true in Test IX-4. In Test X-2, the first record of (+ − − + +) would have been plus if there had been *one* more success which he did get 3 months later. In other words, he receives 2 months credit for detecting *one* more absurdity on the retest than he had on the previous test. In the Reading and Report Test Year X-4, success requires that the selection be read in 35 seconds allowing 2 mistakes in reading and that 8 "memories" be reported. On Jack's first test his record was 14 memories, 30 seconds, and 3 mistakes. The 3 mistakes in reading were fatal or at least the third one was, for he is allowed only 2. Now, if he can manage 3 months later to make only 2 mistakes in reading his mental age will go up 2 months. Does he do it? Fourteen memories, 25 seconds and no mistakes in reading. And for this improvement, after 3 months of school training, he gets 2 months credit toward his mental age.

On the other hand note the tests where on the re-examination he

either did more poorly than he did on his first test or else made mistakes different from the ones made before. Note IX-1, IX-3, X-5, X-6.

The 9-months credit earned on VIII-1, IX-2, IX-4, XII-7 we do not begrudge Jack, for they stand for either partial or complete success on the second test as against *absolute failure* on the first. Tests VIII-3, X-2, X-4, however, represent rather questionable gain since they are earned through such slight increments over partial successes.

For the sole purpose of an interesting if entirely unwarranted and unscientific experiment, let us see what happens to Jack's IQ if we add these last named and questionable 6 months credit to his Mental Age on the first examination, as though they had been successes instead of only partial successes. This would give an IQ of 88 on the first examination as against an IQ of 90 on the second. (As a matter of fact, this is probably really nearer the truth than the unaltered data unless we put intelligent interpretation on these data.) Would this much difference in IQs seem to indicate that the IQ is worthless or would it hint that Jack is a little below "average" but near the upper end of the "dulls?" This seems to be supported by several facts, namely that he has the vocabulary of only an 8-year-old and that in 3 months he increased that vocabulary score by only 1; by the fact that his basal year was as low as 7 considering that he is a 10-year-old child; and by the character of some of his responses.

And now to turn to the case of Mary. We see that she gained 15 months and lost 3 on her re-test, making an apparent gain of 12 months in exactly 4 months to the day. By inspection of her chart, however, we find that the XII year Ball and Field success on the second test is absolutely the only plus score where there had been an *entire* failure on the first test. In fact, the marginal failures in the first test are very conspicuous when compared with the corresponding successes on the second test: IX-1, IX-2, IX-5, X-4, X-5, and X-6. Also as in the case of Jack, we find here some partial failures on the second test where there had been none before and also one final minus score as against a plus (XII-8).

Now that we have Mary's gain accounted for, what is to be said about her *real* IQ? Is it nearer 128 or 119? If an IQ of 120 is to be considered a dividing line (Terman 110–120 "superior;" 120–140 "very superior") to which group does the child really belong? With an IQ on the first test only one point below the lower limits of "very superior" and containing 5 tests totalling 10 months credit in each of which one single more success in addition to the ones already earned

would have been rewarded with a "plus" score, one sees the necessity for a critical evaluation of this 119 IQ. Certainly this child is far superior to a child with an IQ of 119 or even 120 who fails *entirely* on such tests as IX-1, X-4. (Note Mary's minuses on these tests, composed of some plus and some minus scores.) Furthermore there was every indication throughout the test administered by the writer of Mary's great superiority as, for example, her self-criticism, choice of words and the like. While the vocabulary is not high and the child does not talk a great deal (as does Jack) yet she "delivers the goods." She is undoubtedly a "*very superior*" child despite the IQ of 119. Fortunately in this case we have one other line of evidence upon which to fall back and that is a test made by still another examiner about a year before the first of our two tests. Without comparing procedure, scoring, etc., it is interesting just to note in passing that the IQ on that test was 125—"very superior."

In conclusion the writer wishes to state as an opinion formed after a considerable testing experience (although the data cited above do not present all the evidence) that:

1. In order to be of any value whatever every IQ must be critically evaluated for the purpose of seeing what factors have contributed to it or militated against it and of seeing what "marginal" failures and successes there may be.

2. Slight variations in IQs from different testings are to be expected and may mean very little. In every case comparisons of IQs should be made in the light of (1) above.

3. In the main the IQ of children up to about 16 years of age (probably excluding the lowest grades of defectives who soon find themselves in institutions) is reasonably constant and reliable and is highly valuable for diagnostic purposes.

NOTES ON ARTICLES IN EDUCATIONAL PSYCHOLOGY IN CURRENT ISSUES OF OTHER MAGAZINES

REPORTED BY CECILE COLLOTON

Department of Educational Psychology, The Lincoln School of Teachers College

Intelligence Tests

Intelligence as Related to Nationality. Gilbert L. Brown. Journal of Educational Research, 1922, April, 324–327. Nine hundred and thirteen children of foreign parentage tested by the Stanford-Binet show wide range of intelligence Germanic groups—Norwegian, German, Swede, English and Austrian test higher than any of the non-Germanic groups.

The Relative Progress of VII-B Groups Sectioned on the Basis of Ability. W. W. Theisen. Journal of Educational Research, 1922, April, 295–305. Advantages of grouping pupils on basis of ability. Needed changes in curricula, achievement standards, and in supervision.

Intelligence Tests and the Classroom Teacher. Arthur W. Kallom. Journal of Educational Research, 1922, May, 389–399. Practical suggestions for the use of intelligence tests by class-room teachers, illustrated by actual data and individual cases.

Some Pitfalls in the Administrative Use of Intelligence Tests. M. R. Trabue. Journal of Educational Research, 1922, June, 1–11. Cites errors made when tests are used by administrators who have no special training or experience in the field of measurement. Shows need for psychologist and administrator to work hand in hand.

How Much Mental Ability Does a Teacher Need? W. B. Bliss. Journal of Educational Psychology, 1922, June, 33–41. Data on relation between mental ratings and success in teaching.

Comparison of the Binet-Simon and Otis Tests. S. C. Garrison and J. S. Tippet. Journal of Educational Research, 1922, June, 42–48. A study of 158 pupils of the Peabody Demonstration School. Six tables present data.

Psychological Examination of Pre-school Age Children. David Mitchell. School and Society, 1922, May 20, 561–568. A study of 1113 pre-school age children of New York City shows the value of a preliminary examination and the need for special classification and modified curriculum. The examinations were conducted by the New York State Association of Consulting Psychologists.

Intelligence Tests and Collegiate Selection. Dagny Sunne. 1922, May 27, 593–595. Experiments with various tests at Newcomb College show that intelligence tests do differentiate students rather definitely, especially in the freshman year.

Mental Tests and College Teaching. Wm. R. Wilson. School and Society, 1922, June 10, 629–635. Discusses lack of co-relation between ability of college students, as shown by mental tests and their college grades. Suggests ways of

making each student work up to the limit of his powers—especially those of the superior group.

The Normal Curve and the Distribution of Intelligence Ratings. Garry C. Myers. School and Society, 1922. June 17, 676–678. Shows distribution in curve form of 5115 scores, Grade I to college inclusive, on Myers Mental Measure intelligence scale.

Sectional Differences as Shown by Academic Ratings and Army Tests. Martha McLear. School and Society, 1922. June 17, 676–678. A study of Northern and Southern negroes on basis of academic standing and army tests. Negligible difference in academic standing. Wide difference in mentality. Northern negroes higher.

Note on a Method for Studying Causes of Increase in Alpha Scores. Margaret V. Cobb and H. A. Tape. School and Society, 1922, June 24, 706–708. A study of the increase in Alpha scores of 81 pupils on three successive examinations comparing the increase from one class to the next, and the increase from year to year of each class group.

The Results of the Thorndike Intelligence Examination in the Senior Class of the Horace Mann School for Girls. Clara F. Chassell. School and Society, 1922, May 6, 511–512. Good evidence as to the fitness of Horace Mann graduates to do college work. The range of scores of the 54 students tested was from 41 to 94. Sixty-six per cent qualified absolutely, 28 per cent can probably do college work if specially industrious. Six per cent would be likely to prove unsuitable material for colleges and universities of high standards.

Eliminating First-grade Failures through the Control of Intellectual, Physical and Emotional Factors. Grace Arthur. School and Society, 1922, Apr. 29, 474–489. A study of 36 Grade I children and their progress during the first year of school. Case studies describe the actual work done in the ungraded room.

The General Philosophy of Grading and Promotion in Relation to Intelligence Testing. Henry W. Holmes. School and Society, 1922, Apr. 29, 457–461. Points out the need for segregation of gifted children and an enriched curriculum for them. Discusses the dangers of rapid advancement.

Shall We Classify Pupils by Intelligence Tests? Frederick S. Breed. School and Society, 1922, Apr. 15, 406–409. The difficulties involved in classification by intelligence tests. Discussion and practical suggestion.

The Discriminative Value of the Sub-tests of a Group Intelligence Test. Dora K. Mohlman. School and Society, 1922, Apr. 8, 399–400. Study of the results secured by administering the Indiana Group Scale of Intelligence to 77 university juniors and seniors. Correlation coefficients are reported for each part of the scale with the total score.

Intelligence as a Factor in the Election of High School Subjects. S. R. Powers. The School Review, 1922, June, 452–455. A comparison of the subjects elected by the high school students of Fort Smith, Arkansas and their scores on the Otis Intelligence Test. Students with high scores choose subjects making largest intellectual demands.

Intelligence Tests as a Basis for Homogeneous Grouping. May M. Harper. Elementary School Journal, 1922, June, 781–782. An experiment in the McKinley Junior High School, Xenia, Ohio, proves that classification on the basis of group intelligence test scores does place pupils fairly accurately.

How Different Mental Tests Agree in Rating Children. W. S. Guiler. Elementary School Journal, 1922, June, 734–744. A comparison of the Stanford-Binet, National Intelligence, Illinois Examination, and Pintner Non-Language Tests on the basis of IQ. Results of this study are invalidated by the fact that IQs computed on other tests are not comparable to the Binet IQ.

Educational Determinism, or Democracy and the IQ. Wm. C. Bagley. Educational Administration and Supervision, 1922, May, 257–272. An attack upon intelligence testing as inimical to the ideals and purposes of democratic education.

Intelligence and Behavior. A. A. Roback. Psychological Review, 1922, January, 54–62. A criticism of the behavioristic interpretation of intelligence. Reference to the definitions of intelligence in Intelligence and Its Measurement—Symposium. Journal of Educational Psychology, 1921, Vol. XII, p. 124.

A Comparison of Mental Abilities of Mixed- and Full-blood Indians on a Basis of Education. Thomas R. Garth. 1922, May, 221–236. Under conditions of uncontrolled social status but controlled school training the mixed-blood Indians surpass the full-bloods. Details and discussion of tests used.

Mental Tests and Mentality. T. H. Pear. Psyche, 1922, April, 304–314. Raises questions as to reliability of mental tests. Suggests paying more attention to "mental apparatus" and "mental attitude."

Intelligence Tests for Prospective Freshmen. Walter Dill Scott. Chicago School Journal, 1922, May, 321–324. A plea for vocational and educational guidance in college through an adequate Personnel Department.

The Revised and Extended Binet-Simon Tests, Applied to the Japanese Children. Y. Kubo. Pedagogical Seminary, 1922, June, 187–194. A revision of the Binet-Simon scale adapted to Japanese children. Performance tests and parts of the Otis and Army tests have been included. Range of ages—2 to 14 years.

Educational Tests

Instruments for Measuring Disciplinary Values of Studies. E. L. Thorndike. Journal of Educational Research, 1922, April, 269–279. A description of a test, made up in two series, *A* and *B*, for measuring the general improvement in generalization, relating, selection and organization as produced in a pupil by the study of grammar, languages, or mathematics.

Scales for Measuring Results of Physics Teaching. Harold L. Camp. Journal of Educational Research, 1922, May, 400–405. Description of the development of three scales designed to measure ability in (1) mechanics, (2) heat, (3) electricity and magnetism. Illustrative exercises and tentative norms are given.

Convenience and Uniformity in Reporting Norms for School Tests. J. Crosby Chapman. Journal of Educational Research, 1922, May, 406–420. Presents a scheme of reporting test scores by setting up nine equally separated levels of achievement for each grade. Criticises McCall's *T.* scale as a uniform procedure.

A Study of Reading and Spelling with Special Reference to Disability. Arthur I. Gates. Journal of Educational Research, 1922, June, 12, 24. Details of a study of the reading and spelling abilities of 135 children. Discussion of various defects associated with disability in these subjects. Brief mention of remedial treatment.

The Efficiency Quotient as a Measure of Achievement. T. L. Torgerson. Journal

of Educational Research, 1922, June 25–32. Advocates the use of an achievement quotient found by dividing the pupil's point score by the grade standard, and an efficiency quotient found by dividing his achievement quotient by his intelligence quotient. Tables for quotients embracing all standard tests have been prepared.

Measuring the Pupils in a Large City School. Joseph S. Taylor. School and Society, 1922, July 1, 25–28. Report of tests conducted by Dr. McCall and a group of graduate students at Public School 107, New York City. Grades tested were III*A* to VI*A*, including an ungraded open-air class and a class of mental defectives. Details of the various scores computed—educational age, pedagogical rank, promotion quotient, achievement quotient, etc., are given.

A Simplified Method of Determining a Pupil's Score on Gray's Oral Reading Test. W. S. Monroe. School and Society, 1922, May 13, 538–539. Using the same zero point for all grades and making the interval between paragraphs 4 instead of 5 provides a simple method of computing and interpreting scores on the Gray test. Norms expressed in terms of the new scale are given.

A Geography Test for the Sixth, Seventh and Eighth Grades. C. A. Gregory and Peter L. Spencer. School and Society, 1922, April 22, 452–456. A complete description of a comprehensive geography test, including discussion of criteria for choice of questions, and directions for giving and scoring the test. There are three duplicate forms each consisting of six parts.

Forecasting Failures in College Classes. Harvey B. Lemon. The School Review, 1922, May, 382–387. Description of a brief test used in the administration of undergraduate work in general physics at the University of Chicago. Elimination of probable failures and establishment of sympathetic acquaintance two of the beneficial results noted.

The Accomplishment Quotient—Finding and Using It. Katherine Murdoch. Teachers College Record, 1922, May, 229–239. A study of 415 children, Grades III to VIII in a large private school in Honolulu, Hawaii. Tests used were National Intelligence, Thorndike-McCall Reading, and Woody-McCall Mixed Fundamentals. Full discussion of process of obtaining accomplishment quotients and suggestions for the use of results.

The Cleveland Survey Arithmetic Test in Grade V-*B in Chicago.* Edw. E. Keener. Chicago School Journal, 1922, May, 336–344. Complete data secured from the use of the Cleveland test in 266 schools. Discussion of standardized tests as aids in teaching.

Miscellaneous

An Aid to the Analysis of Vocational Interests. J. B. Miner. Journal of Educational Research, 1922, April, 311–323. Description of an individual analysis blank which aims to train pupils to analyze their work interests. Two new features are a classification of occupations according to activities emphasized and a presentation of contrasts in working conditions.

An Accurate Index of Nationality. Riverda H. Jordan. Journal of Educational Research, 1922, May, 421–425. Argues that in studying the nationality of school children data should be based upon the birthplaces of grandparents rather than fathers.

The Psychology of Learning Applied to Typewriting. E. W. Barnhart. The American Shorthand Teacher, 1921, November, 1922, February. Four articles on the application of the well-established laws of learning to typewriting.

A Data Sheet for the Pearson Correlation Coefficient. L. L. Thurstone. Journal of Educational Research, 1922, June, 49–56. A labor-saving device for the calculation of the Pearson *r*. Illustration of data sheet and complete instructions for its use.

A Critique of Mental Measurements. Thomas J. McCormack. School and Society, 1922, June 24, 686–692. Expresses complete lack of belief in the whole science of mental measurement.

A New IQ Slide Rule. Lloyd N. Yepsew. School and Society, 1922, May 27, 596. Description of a circular rule device designed especially for use in computing intelligence quotients.

Classification in Athletics for the Purpose of Individual Self-rating. Jesse Feiring Williams and Myrtle Hummer. Teachers College Record, 1922, May, 240–254. Class divisions based on the actual accomplishment records of 1612 boys and 1773 girls of Trenton, N. J., permit self-rating in athletics and physical ability. Eleven interesting charts present details.

Variation in Grading High School Pupils. J. E. Armstrong. Chicago School Journal, 1922, May, 346–348. A criticism of our present grading system.

Vocational Interests of High School Seniors. Aubrey A. Douglass. School and Society, 1922, July 15. Analysis of the answers of 1658 girls and 1186 boys—all high-school seniors in the State of Washington—to a vocational interests questionnaire.

NEW PUBLICATIONS IN EDUCATIONAL PSYCHOLOGY AND RELATED FIELDS OF EDUCATION

1. *Mental Growth Curve of Normal and Superior Children Studied by Means of Consecutive Intelligence Examinations.*[1]

Dr. Baldwin was inspired with a fine idea when he conceived the plan of measuring at successive intervals the intelligence of a fairly large number of children. The Stanford-Binet examination was given to 56 subjects twice, to 51 three times, to 44 four times and to 36 subjects five times. In all 143 individual record cards were used. The excellent plan of computing the mental age for each exact chronological age was also followed. According to the authors (1) "The mental growth curve reveals a significant change in the trend with the approach of adolescence, which appears earlier in the case of superior children. There is also an adolescent superiority of girls which is in accordance with other facts indicative of the earlier maturity of girls;" (2) "the mental growth curves are strikingly similar to the physical growth curves in height;" (3) "the IQ curves are approximately horizontal, confirming within limitations the constancy of the IQ. There are fluctuations associated with physical development;" and (4) "the mean IQ of each of the four groups of children increased with each successive examination, which is probably an effect of greater habituation or practice."

So far, so good! If one accepts Dr. Baldwin's method of plotting the curves one can accept his conclusions. But hasn't he made an error in logic in this part of his task? To show mental growth he plots mental ages against chronological ages, and apparently is astonished when both the superior and the inferior group curves are roughly straight lines. What else could be expected? His method of plotting (Chart II) simply proves what other studies have also proved, namely, that the IQ from 5 to 14 is fairly constant. That this criticism is not mere quibbling is shown by comparing Chart II (a mental growth curve) with Chart IV (an IQ curve). They are the same thing

[1] Baldwin, Bird T., and Stecher, Lorle I.: University of Iowa Studies, Vol. II, No. 1, January, 1922, pp. 61.

plotted in different fashion, although diametrically opposite conclusions are drawn from them. (See conclusions (1) and (3) above.)

Isn't Dr. Baldwin thinking in a circle when he uses mental age as a measure of mental growth? Isn't he assuming that the mental growth from 5 to 6 is the same as the mental growth from 12 to 13? It may be that it is so, but no one has yet proved it. Until an absolute unit for measuring intelligence is devised it will be impossible to say that the curve of growth of mental age is similar to the curve of growth for height, or to any other curve for that matter. Theoretically, the curve should be logarithmic. And the records from a large number of group tests give indications of a logarithmic character. But until the absolute unit has been discovered such researches as this are quite beside the mark. What Dr. Baldwin has done is a good piece of work on the question of the constancy of the IQ. He shows conclusively that it is fairly constant though subject to fluctuations so far as the measurements made are reliable. He has also pointed out that the results of repeated examinations exhibit a definite practice effect. But he has *not* plotted a mental growth curve although it pains the reviewer to have to point this out. But if it will soothe Dr. Baldwin's feelings the writer confesses that he (as well as many others) has previously stumbled into this very error.

PETER SANDIFORD.

---

2. *A Study of Superior Children.*—The gradually increasing number of studies dealing with superior children shows that psychologists and educators are beginning to realize the importance of knowing more about this type of child. The author of the monograph under consideration[1] presents in sociological and psychological study of a small group of children of superior intelligence. The children were selected from among those reported by principals and teachers as superior. The younger children had IQs of 135 or above, and the older had IQs of 120 or above. One or two cases with lower IQs were studied, the lowest being 117. These children were given a great number of tests of all kinds, opposites, symbol-digit, directions, proverbs, and the like. The superior children excelled normal children on all these tests.

The sociological part of the study includes very elaborate case histories of each child. These are very interesting in as much as we

---

[1] Root, W. T.: A Socio-Psychological Study of 53 Supernormal Children. *Psychological Monographs*, Vol. 29, No. 4. Whole No. 133. Princeton, 1921.

have little of this sort of data as compared with the amount we possess for subnormals. In general the superior child is characterized as having a good home and superior parents. The author then attempts to explain his results and proceeds to a long discussion of the central common factor theory. According to our author the common factor could just as well be environmental as innate. "The common factor or factors may be a varying admixture of innate ability, formal training, incidental education and social conditions." Indeed, all through the thesis the author emphasises the importance of environmental factors much more than the average psychologist would, and one is inclined to question the soundness of some of his opinions in this respect. It is, however, well to have this side of the picture presented, even although we cannot agree. There seems, furthermore, a feeling of dissatisfaction on the part of the author with the present intelligence tests or with tests in general. He feels that no tests measure the real basis of superior intelligence, for this real basis consists of ability to suspend judgment, freedom from suggestibility, critical attitude, etc. Again there is room for much argument over these phrases. The results presented in the monograph are important and interesting. The conclusions and the opinions of the author are open to much debate.

R. P.

---

3. *Group Intelligence Tests in England.*[1]—This book gives American psychologists the first account of group testing in England and we have in its author, Mr. Ballard, a most delightful cicerone for our tour, one who knows well how to mix humor with his learning. We must remember, however, that he is not acting as guide to Americans, but is explaining the field to English teachers. So much the more interesting, therefore, is it to the American psychologist to hear the explanation of our own group tests. Here is their origin. "Individual testing was born in France; group testing was born in America. And its mother was necessity—the stern necessity of war." And so he tells of the testing in the army. "The whole undertaking was a colossal business; and the official report which has recently been issued is correspondingly colossal. It weighs about four pounds." Mr.

[1] Ballard, P. B.: Group Tests of Intelligence. Hodder and Stoughton, London, 1922, pp. X, 252.

Ballard admires our fool-proof tests, and he says very aptly that "a fool-proof test is one that prevents the examiner from making a fool of himself." And so he ambles along with a delightfully humorous, but thoroughly sympathetic, account of the work in this country. It is an account that no psychologist should fail to read.

For American readers the description of group testing in England is of particular interest. Although little has so far been done, the use made of intelligence tests by the Bradford and Northumberland Education Committees for the selection of scholarship children is important and significant, and one feels that in the course of time the extent of the work must inevitably increase. Our author also gives samples of several group tests used in England. Many of these are clearly descendents of the American group test, but there are innovations and new ideas that are very suggestive. The English workers are not quite as partial to short time limits as most American workers. New types of material for group tests are Thomson's Hindustani Test and Ballard's Cipher Test, Orientation Test and Cryptogram Test.

The book also contains a very simple chapter on the nature of intelligence in which the various definitions of intelligence are discussed. It is admirably written and forms a decided addition to the literature of the subject. It ends with the author's own definition as follows, "Intelligence is the relative general efficiency of minds measured under similar conditions of knowledge, interest and habituation."

Other chapters in the book deal in a simple manner with correlation and statistical procedure, and with the use of group tests in schools. The last chapter in the book has no connection with the book proper. It is on "Spelling Demons." The reviewer wonders why the author included this unrelated topic. Perhaps as a mental test for reviewers to see if they could discover the relationship of this small part to the whole book. If so, the present reviewer has failed.

R. P.

---

4. *Tests of College Students.*—The college student has been the most popular "research animal" in the psychological laboratory, and the present monograph[1] is another addition to the long list

[1] Carothers, F. E.: Psychological Examinations of College Students. *Archives of Psychology*, No. 46, December, 1921.

of studies reporting psychological tests of college students. It contains a very good summary of the previous work done in this connection. The contribution of the author herself is a detailed study of 19 different tests given to various groups of college freshmen. The results show the possibility of finding "several groups of tests which correlate closely among themselves, but loosely with the other tests." There is no evidence in support of Spearman's theory of a general common factor in so far as this depends upon the inter-correlations.

The author's attempt to explain the correlation, or the lack of it, of the separate tests with various college subjects is very naive. It is hard from a study of the actual coefficients to believe, as the author tells us, that "the five academic groups show positive correlation with tests which we would expect to correlate with them." Mathematics shows the highest correlations with Cancellation, Checking, and Knox Cube; Science with Opposites, Verb-objects, Mixed Relations, Knox Cube, and Logical Recollection. An attempt to explain why is made by the author on the basis of the mental processes involved. Fortunately she refrains from telling us why Philosophy should correlate highest with Cancellation, Word Naming, Knox Cube, and Digit Span.

The practical value of psychological tests for students' guidance is stressed. A psychographic chart for each student is recommended, which will show respective strengths and weaknesses.

The work is carefully done and is a valuable addition to our knowledge of the psychology of the college freshman. R. P.

---

5. *Juvenile Delinquency.*—The book under consideration[1] presents in interesting form a report of the work being carried on at The Ohio Bureau of Juvenile Research. As such it is of value to educators, who have to deal with disciplinary problems, and with the questions arising from the presence of mental deviates in schools.

A special chapter is devoted to the psychopathic child, who may be of any degree of intelligence, but who shows abnormalities of mental functioning apart from capacity for learning. A state institution for psychopathic children is recommended.

The author believes that "juvenile delinquency can be largely

[1] Goddard, H. H.: "Juvenile Delinquency." Dodd, Mead and Company, New York, 1921, pp. 120.

eradicated" if the states are willing to pay the price of studying and caring for unfortunate children. Research should be one of the state's chief concerns in the present condition of knowledge, for as yet many fundamental questions relating to the origin, development and modifiability of unfortunate deviates cannot be answered.

LETA S. HOLLINGWORTH.